



JOURNAL

Submitted (16-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
On Jul 12, 2001 this sequence version replaced gi:12654174.  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalobom.tmc.edu](mailto:villalobom.tmc.edu)  
Villalobon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Murthy, D.M., Gibbs, R.A.

FEATURES

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 3 Row: k Column: 21  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4886442.  
Location/Qualifiers  
1..1433

source

CDS

```

/organism="Homo sapiens"
/db_xref="LocustID:5861"
/db_xref="taxon:9606"
/clone="MGC:5233 IMAGE:290705"
/tissue_type="Placenta, Chorioncarcinoma"
/clone_11b="NH_MGC_10"
/lab_host="DH10B"
/Note="Vector: pCMV-SPORT6"
179..796
/codon_start=1
/product="RAB1, member RAS oncogene family"
/protein_id="AAH00905.1"
/db_xref="gi:12654175"
/translation="MSSMNPEDYLFKLLIGDSGVKSCLLRFPADPTYESYISTIGVDFRIEIDGKTIKLIOWDTAGOEPRITSSYVAGHAIYVYVTOESNNYKOWLEIDRVASNNKLLNGKIDLTETTKKYVDYTAEPHPSLGLPIPLEHSARNAIVVQSFMTAAETKRKMGATGACGAEKSNVXIOSTPVAOSGGCC"
BASE COUNT      432 a      254 c      326 g      421 t
ORIGIN

```

Query Match 86.8%; Score 1219; DH 9; Length 1433;  
Best Local Similarity 93.2%; Pred. No. 2.8e-187;  
Matches 1336; Conservative 0; Mismatches 0; Indels 97; Gaps 2;

```

3 GCGATAGCTGAGT -GCGGCGGCTGCTGATTTGTTCTAGAGGACGAGTAGGGAAGACG 61
1 GCGATAGCTGAGTGGCGGCGGCGGCTGATTTGTTCTAGAGGACGAGTAGGGAAGACG 60
62 TTTGCTCTCCGGAACAGCTTATTCATTCCTTTCTTCGATTACCGCTGGCGGAGAG 121
61 TTTGCTCTCCGGAACAGCTTATTCATTCCTTTCTTCGATTACCGCTGGCGGAGAG 120
122 TCAGGCGGCGGCTGGCGGACAGAGGCGGCGGCTGGCGGCGGCGGCGGCGGCGGAGAT 181
121 TCAGGCGGCGGCTGGCGGACAGAGGCGGCGGCTGGCGGCGGCGGCGGCGGCGGAGAT 180
182 GTCCAGCATGATCCCGAATATGATTTATTTCAAGTACTTTCGATTGGGACTCAGG 241
181 GTCCAGCATGATCCCGAATATGATTTATTTCAAGTACTTTCGATTGGGACTCAGG 240
242 GGTGGGAAGTCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 301
241 GGTGGGAAGTCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 300
302 CAGCACAATTTGCTGATTTCAAAATAGAACTTATAGTTAGCGGGAACCAATCAA 361
|||||

```

```

301 CAGCACAATTTGCTGATTTCAAAATAGAACTTATAGTTAGCGGGAACCAATCAA 360
362 GCTTCAATA----- 371
361 GCTTCAATAATGAGACACAGCAGCGGAGAAAGATTGGAACATCACTCCATCTTATTA 420
372 -----GAGTCTTCAATA 385
421 CAGAGAGCCCATGCGCATATAGTTGTATGATGTGACAGTACGAGTCCCTCAATTA 480
386 TGTTAACAGTGGCGCAGAGAAATGATCGTTATGCCGATGAAATGTCACCAATTTGT 445
481 TGTTAACAGTGGCGCAGAGAAATGATCGTTATGCCGATGAAATGTCACCAATTTGT 540
446 GTTAGGAACAAATGATGATGACACAAAGAAAGTATGACTACATACACAGGAGGA 505
541 GTTAGGAACAAATGATGATGACACAAAGAAAGTATGACTACATACACAGGAGGA 600
506 ATTTGCTGATTCCTTGGAAATTCGTTTGGAAACCGAGTCTAAGATGCAAGATGT 565
601 ATTTGCTGATTCCTTGGAAATTCGTTTGGAAACCGAGTCTAAGATGCAAGATGT 660
566 AGAAGAGCTTTTCATGAGAGAGAGCGAGCTGATTAAGAGGATGGTCCCGAGACAC 625
661 AGAAGAGCTTTTCATGAGAGAGAGCGAGCTGATTAAGAGGATGGTCCCGAGACAC 720
626 AGCTGGTGGTCTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGAGTCA 685
721 AGCTGGTGGTCTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGAGTCA 780
686 TGGAGGTTGCTGCTTAAATTTGGCTCCATTCCTTTTCACAGCAATGATTTGCAAT 745
781 TGGAGGTTGCTGCTTAAATTTGGCTCCATTCCTTTTCACAGCAATGATTTGCAAT 840
746 AACCAGATGAAAAAACAATTTGCTTAAATTTGCTTAAATTTGCTTAAATTTGCTTAA 805
841 AACCAGATGAAAAAACAATTTGCTTAAATTTGCTTAAATTTGCTTAAATTTGCTTAA 900
806 TTCTTACGCTCCCAAGAGTCCAGAGATTTGTAATGCTCAATCTGATTTTAT 865
901 TTCTTACGCTCCCAAGAGTCCAGAGATTTGTAATGCTCAATCTGATTTTAT 960
866 TCCCTTGCTCAAGACAGCTTAACTTCAATTTCAAGACTGTTTAAACCTTTGCTGCT 925
961 TCCCTTGCTCAAGACAGCTTAACTTCAATTTCAAGACTGTTTAAACCTTTGCTGCT 1020
926 TTTTAAATATATGATGATTCCTTGTGCTTCCGATTCAGAGCTTCCGCT 985
1021 TTTTAAATATATGATGATTCCTTGTGCTTCCGATTCAGAGCTTCCGCT 1080
986 TGGTAAATATATTTGTTTGAATGTTTATTTGCAATTTGCAATTTGCAATTTGCA 1045
1081 TGGTAAATATATTTGTTTGAATGTTTATTTGCAATTTGCAATTTGCAATTTGCA 1140
1046 TTTGGAAGATGAGTTCAGCCATTTGTAATCAACAGCAGCAGAGTGTGCTACATTC 1105
1141 TTTGGAAGATGAGTTCAGCCATTTGTAATCAACAGCAGCAGAGTGTGCTACATTC 1200
1106 CATGATTAAGTTAGTGAAGATTTATTTGAATGATTTGATTTGATTTGATTTGAT 1165
1201 CATGATTAAGTTAGTGAAGATTTATTTGAATGATTTGATTTGATTTGATTTGAT 1260
1166 GAGTTAATTAATGAAAGATTTACATATGATTTATTTGATTTGATTTGATTTGAT 1225
1261 GAGTTAATTAATGAAAGATTTACATATGATTTATTTGATTTGATTTGATTTGAT 1320
1321 TTTGGCTGAGAGATTTGTAATTTGTTGACACTATGTAACAAACAACTGAAGATAT 1380
1286 GTTTAATTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1338
1381 GTTTAATTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1433

```



Query Match	74.9%	Score 1052.6	DB 10	Length 1444
BASE COUNT	412 a	270 c	342 g	420 t
ORIGIN				
FEATURES	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a></p> <p>Series: IRAK Plate: 11 Row: a Column: 2</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6679586.</p> <p>Location/Qualifiers</p> <p>1..1444</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/map="FVB/N"</p> <p>/clone="MGC:6226 IMAGE:3592802"</p> <p>/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."</p> <p>/clone_lib="NCI CGAP_Maml"</p> <p>/lab_host="DH10B"</p> <p>/note="vector: pCMV-SPORT6"</p> <p>196..813</p> <p>/codon_start=1</p> <p>/product="RAB1, member RAS oncogene family"</p> <p>/protein_id="AAH02077.1"</p> <p>/db_xref="GI:12805233"</p> <p>/db_xref="LOCUSID:5861"</p> <p>/translation="WSSNPEPYDYLFLLLLIGDSGVKSCILLRFADDTYESISTTGVDVRIETDELDTKTLKQLWDAGQERFRITSSYRGAGIIVYDVDOESFNWVGMQKQIEDIRASENVNKLWGNCMDLTKTKVDYTKAFEDSLGIFLETASANNANVVEQSFMAAEIKRKRMGPAGATAGCAKSNVXIISTPVKSGGGCC"</p>			
REMARK	<p>NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a></p> <p>Contact: MGC help desk</p> <p>Email: <a href="mailto:cgagbs-remail.nih.gov">cgagbs-remail.nih.gov</a></p> <p>Tissue Procurement: Gilbert Smith, Ph.D.</p> <p>cDNA Library Preparation: Life Technologies, Inc.</p> <p>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)</p> <p>DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305</p> <p>Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a></p> <p>Contact: (Dickson, Mark) <a href="mailto:mcdspaxil@stanford.edu">mcdspaxil@stanford.edu</a></p> <p>Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.</p>			
COMMENT				
REFERENCE	<p>1 (bases 1 to 1444)</p> <p>Strausberg, R.</p> <p>Direct Submission</p> <p>Submitted (31-Jan-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA</p>			
AUTHORS				
TITLE				
JOURNAL				
ORGANISM	<p>house mouse.</p> <p>Mus musculus</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.</p>			
DEFINITION	<p>BC002077 1444 bp mRNA linear ROD 07-AUG-2002</p> <p>Mus musculus, RAB1, member RAS oncogene family, clone MGC:6226 IMAGE:3592802, mRNA, complete cds.</p>			
ACCESSION	BC002077			
VERSION	BC002077.1			
KEYWORDS	GI:12805232			
SOURCE	MGC.			
LOCUS	BC002077			
DB	1178			
QY	1320 AAAA 1323			
DB	1238 CAAA 1241			
RESULT 3				
BC002077				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
DEFINITION				
ACCESSION</				

Best Local Similarity 88.3%; Pred. No. 1,9e-160;		
Matches 1256; Conservative 0; Mismatches 59; Indels 108; Gaps 6		
QY	16	GCGGCGCGCTGATTTGTGTTCTAGAGGAGCAGAGTAGGAGGAAGCGTTTGCCTCCGGA 75
Db	27	GCGGCGCGCTGATTTGTGTTCTAAGGATCGGAGT -GGAGTGCAGTTTGTCTACCGGA 85
QY	76	ACAGCCTATCATTTCCCTTCCTTCGATTAACCCCGGGGGGAGAGTCAAGGGCGGGCT 135
Db	86	ACAGTTTAGCTATCTCTCCCTTCCTTCACTTACCTGTGGCGGAGATTGGGGCGGGCT 145
QY	136	GCGGCAGCAGAGGCG -GGCGGTGGCGGGCGGGCGGAGCTGCAGTGCATGTCTCAGCA 189
Db	146	GCCTCAGCAAGGCGGTGGTGGCGGGCGGGCGGGGCGGAGCTGCAGTGCATGTCTCAGCA 205
-QY	190	TGAATCCCGAATATGATTTATTTATTCAGATTACTTCTGATTTGGCAGCTCAGGGGTTGGAA 249
Db	206	TGAATCCCGAATATGATTTATTTATTCAGATTACTTCTGATTTGGCAGCTTCTGGGGTTGGAA 265
QY	250	AGCTTGCGCTCTCTTCTTGTAGTTTGCAGATGCATATATACAGAAAGCTATACAGACACA 309
Db	266	AGTCTGCGCTCTCTCTTGTAGTTTGCAGATGCATATATACAGAAAGCTATACAGACACA 325
QY	310	TTGTGTGTGATTTCAAATAAAGAACTATAGATTAGACGGGAAACAATCAAGCTTCAAA 369
Db	326	TTGTGTGTGATTTCAAAGATACGAACCTATAGATTAGATGGGAAACAATCAAGCTACAGA 385
QY	370	TA----- 371
Db	386	TATGGGACACAGCAGGCCAGGAAGATTCCAAACATACCTCCAGTTATTACAGAGNG 445
QY	372	-----GAGTCTTCAATAATGTTAAAC 393
Db	446	CCCATGGCATATAGTTGTGTATGATGTGACAGATCAGAGAGTCCTTCAATAACGTTAAAC 505
QY	394	AGTGGCTGCAGGAATATGATGTGATGCATGAAATGTCAACAATTTGTGTGTAAGGA 453
Db	506	AGTGGCTGCAGGAATATGATGTGATGCATGAAATGTCAACAATTTGTGTGTAAGGA 565
QY	454	ACAAATGTGATCTGACCAACAAAGAAAGTAGTAGACTACACAACAGCGAAGGAATTTGCTG 513
Db	566	ACAAATGTGACCTGACCAACAAAGAAAGTAGTAGACTACACAACAGCAAAAGGAATTTGGAG 625
QY	514	ATTCCTTGGAAATTCCTTTTGGAAACCAAGTGCCTAATAATGGCAACGAATGTAGAAACAGT 573
Db	626	ATTCCTTGGAAATTCATTTTGGAAACCAAGTGCCTAATAACGCCAAGATGTAGAAACAGT 685
QY	574	CTTTCATGACATGGCAGCTGAGATTTAAAGCGAATGGTCCCGGACCAACAGCTGGTG 633
Db	686	CTTTCATGACATGGCAGCTGAGATTTAAAGCGAATGGTCTCGGAGCTACAGCTGGTG 745
QY	634	GTGCTGAGAACTCCAATGTTAAATTCAGACCACTCCAGTCAAGCAGTCCAGGTGGAGGTT 693
Db	746	GTGCGCGAAGATCCAATGTTAAATTCAGACCACTCCAGTCAAGCAGTCCAGGTGGAGGCT 805
QY	694	GTCTCTAAATTTGCTCCATCCCTTTCTCACAGCAATGAGATTTGCATCTGAACCCCAAG 753
Db	806	GTCTCTAAATTTGCTCCATCCCTTTCTCACAGCAATGAGATTTGCATCTGAACCCCAAG 865
QY	754	TGAAAAAACAAATTTGCTCTGAAATGTACTGTATGTAGTGGCAGCTACACAACATTTCTTACC 813
Db	866	TGAAAAAACAAATTTGCTCTGAAATGTACTGTATGTAGTGGCAGCTACACAACATTTCTTACC 925
QY	814	GTCCTCCAAAGGTCAGAGATTGTATATGTGCATACTGAC -TTTTTTTTTATTCCTCTTG 872
Db	926	GTTTTCAC -AAGGTCAGAGATTGTATATATGTGCATACTGACATTTTTTTTTTATTCCTCTTG 984
QY	873	ACTCAGACAGCTTACTTTCATTTTCAGAACTGTTTAAACCTTTGTGCGGTATATAA 932
Db	985	ACTCAGACAGCTTACTTTCATTTTCAGAACTGTTTAAACCTTTGTGCGGTATATAA 1044
QY	933	AATATATGTGTAACTCTTGTGCTTTCGTATACAGACATGTTTCCGTGGTGGTAG 992



Db	1045	AAATATGTGTAACTCTGCTTCTCTGTAACCAAGATCGTTCCCGTGGTTAG	1104
QY	993	AAATATATTTTGTGATGTTATATATGCAATGTTAGATGCAAGTTAGTCTTGAA	1052
Db	1105	AAATATATTTTGTGATGTTATATATGCAATGTTAGATGTTAGTCTTGAA	1164
QY	1053	GATGAGTTGACCAATTTTGTATCAACAGCAAGAGTGTCTGACTTCCATGCAT	1112
Db	1165	GATGAGTTGACCAATTTTGTATCAACAGCAAGAGTGTCTGACTTCCATGCAT	1224
QY	1113	AAATGTTAGTGAAGTTATATATGATGATCTGATTTGCTGCTCTCTGATAGTTAT	1172
Db	1225	AAATGTTAGTGAAGTTATATGATGATCTGATTTGCTGCTCTCTGATAGTTAT	1284
QY	1173	AAATGAGAAATATACATCTGATTAATAGTTCTGATCTGATCTGATTAATAGTTAT	1232
Db	1285	AAATGAGAAATATACATCTGATTAATAGTTCTGATCTGATCTGATTAATAGTTAT	1344
QY	1233	CTGAGAAATATGTTATGTTGTCACACTATGTAACAAACATGATATGTTAT	1292
Db	1345	CTGAGAAATATGTTATGTTGTCACACTATGTAACAAACATGATATGTTAT	1401
QY	1293	AAATATTTGCTTATTTGATGATGATTAATTAATTAATTAATTAATTAATTAAT	1335
Db	1402	AAATATTTGCTTATTTGATGATGATTAATTAATTAATTAATTAATTAATTAAT	1444
RESULT 4			
LOCUS	HSN800023	1208 bp	mRNA linear PRI 10-MAR-2001
DEFINITION	Human sapiens mRNA: CDNA DKFZP564B163 (from clone DKFZP564B163);		
ACCESSION	AL050268		
VERSION	AL050268.1	GI:4886442	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	Wiemann, S., Well, B., Wellenreuther, R., Gassenhuber, J., Glasl, S., Ansoer, M., Boecker, M., Bloeker, H., Bauersachs, S., Blum, H., Lauber, J., Duesterhoeft, A., Beyer, A., Koehler, K., Strack, N., Mewes, H., Ottenwaelder, B., Oberaler, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M., and Poustka, A.		
TITLE	Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs		
JOURNAL	Genome Res. 11 (3), 422-435 (2001)		
FEATURES	2 (bases 1 to 1208)		
REFERENCE	Wambutt, R., Heubner, D., Mewes, H., Gassenhuber, J., and Wiemann, S.		
AUTHORS	Direct Submission		
TITLE	Submitted (10-MAR-1999) WIPs, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY		
JOURNAL	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ). Email: s.wiemann@dkfz-heidelberg.de; Consortium of the German Genome Project. This clone (DKFZP564B163) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY. Email: clonerzpd.de Further information about the clone and the sequencing project is available at <a href="http://www.mips.biochem.mpg.de/proj/cdna/">http://www.mips.biochem.mpg.de/proj/cdna/</a> .		
FEATURES	1..1208		
SOURCE	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="DKFZP564B163"		
	/tissue_type="brain"		
	/clone_id="564 (synonym: hbr2). Vector pAMP1; host		
	X1-2blue; sites NotI + SalI		
	/dev_stage="fetal"		
	191..580		
gene			

CDS	/gene="DKFZP564B163"		
	191..580		
	/gene="DKFZP564B163"		
	/note="Rab1, splice variant"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="CA43369.1"		
	/db_xref="GI:4886443"		
	/db_xref="SPTREMBL:Q9Y372"		
	/translation="MSSMNPEDYLFKLLIDSGVSGKLLRFPADPTYESYISTI		
	GVDFKIRITLDGKTIKIQIEFADSLIGIPIETSAKNATVNDQSFYMAAEIKRMGP		
	GATGAEKSNVYIOSTFVKSGGGCC"		
	1169..1174		
	polya_signal		
	polya_site		
	BASE COUNT 344 a 213 c 279 g 372 t		
	ORIGIN		
	Query Match 74.3% Score 1044; DB 9; Length 1208;		
	Best Local Similarity 90.0%; Pred No 4.7e-159;		
	Matches 1197; Conservative 0; Mismatches 0; Indels 133; Gaps 2;		
QY	1	AAAGCAGTACAGTACGCGCGGCTGCTGATGTTCTTCTGAGGCAAGGAGGAGGA	59
Db	11	AAAGCAGTACAGTACGCGCGGCTGCTGATGTTCTTCTGAGGCAAGGAGGAGGA	70
QY	60	CGTTTCTCTCCGGAACAGCTATCTATCTCTTCTTCTGATTAACCGTGGCGGAG	119
Db	71	CGTTTCTCTCCGGAACAGCTATCTATCTCTTCTTCTGATTAACCGTGGCGGAG	130
QY	120	AGTACAGGCGCGGCTGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGAG	179
Db	131	AGTACAGGCGCGGCTGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGAG	190
QY	180	ATGTCACAGTAAATCCGAAATATATATTTATTCAGTTACTTCTGATGCGCACTCA	239
Db	191	ATGTCACAGTAAATCCGAAATATATATTTATTCAGTTACTTCTGATGCGCACTCA	250
QY	240	GAGGCTTGAAGATCTTCCCTCTTCTTCTGATGTTGCAATGATATATACGAAAGCTAC	299
Db	251	GAGGCTTGAAGATCTTCCCTCTTCTTCTGATGTTGCAATGATATATACGAAAGCTAC	310
QY	300	ATCAGACAAATTTGCTGTGATTTCAATTAAGAATATATATATATATATATATAT	359
Db	311	ATCAGACAAATTTGCTGTGATTTCAATTAAGAATATATATATATATATATATAT	370
QY	360	AAAGCTTCAATATAGTCTTCAATATATATATATATATATATATATATATATAT	419
Db	371	AAAGCTTCAATATAGTCTTCAATATATATATATATATATATATATATATATAT	382
QY	420	GCCAGTAAATATATATATATATATATATATATATATATATATATATATATAT	479
Db	383	-----	382
QY	480	GTAATGACTACACACAGCAAGAAATTTGCTGATTCCTTGAATTCGTTTGGAA	539
Db	383	-----	418
QY	540	ACAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	599
Db	419	ACAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	478
QY	600	AAAGAGCAATGCTGCTCCGAGCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG	659
Db	479	AAAGAGCAATGCTGCTCCGAGCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG	538
QY	660	CAGAGCTCTCAGTCAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	719
Db	539	CAGAGCTCTCAGTCAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	598
QY	720	TCTCAGCAATGATTTTCAATCTGAAACCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	779
Db	599	TCTCAGCAATGATTTTCAATCTGAAACCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	658

OY		780	ACTGATGTACCTGCAGTACAACAAGATTCTTAACCGTCGCCAACAAGGTCAGAGATTGTAA	839
Db		659	ACTGATGTACCTGCAGTACAACAAGATTCTTAACCGTCGCCAACAAGGTCAGAGATTGTAA	718
OY		840	ATGGTCAATACCTGACTTTTATTTTATTTCCCTTGACTCAAGCAGCTAACCTCATTTTCAG	899
Db		719	ATGGTCAATACCTGACTTTTATTTTATTTCCCTTGACTCAAGCAGCTAACCTCATTTTCAG	778
OY		900	AACGTGTTTAAACCTTGTGTGCGGCTTATAAATAATGTGTGAATCCCTGTGCTTT	959
Db		779	AACGTGTTTAAACCTTGTGTGCGGCTTATAAATAATGTGTGAATCCCTGTGCTTT	838
OY		960	CCTGATACACAGACTGTTTTCCCGTGGTGGTTAGAATATATTTTGTGTGATGTTATATTT	1019
Db		839	CCTGATACACAGACTGTTTTCCCGTGGTGGTTAGAATATATTTTGTGTGATGTTATATTT	898
OY		1020	GGCATGTTTAAATGTCAGAGTTTAGCTTCCTGAANAAGTGAAGTTCAGCCATTGTATCAA	1079
Db		899	GGCATGTTTAAATGTCAGAGTTTAGCTTCCTGAANAAGTGAAGTTCAGCCATTGTATCAA	958
OY		1080	CAGCAACAAGCAGTGTCTGCTCACTTCCATGCAATAAGTTTAGTGAGATGTATATGTAA	1139
Db		959	CAGCAACAAGCAGTGTCTGCTCACTTCCATGCAATAAGTTTAGTGAGATGTATATGTAA	1018
OY		1140	ATCTGATTGCTAGTCTTCTCTCTTGACAGTTATAATGGAAGAAAGATTACATCTCATTA	1199
Db		1019	ATCTGATTGCTAGTCTTCTCTCTTGACAGTTATAATGGAAGAAAGATTACATCTCATTA	1078
OY		1200	ATATTTCTTCATACCTGCATATATATATTTGGCGTCGCAAAATATGTATTTGTTGCACA	1259
Db		1079	ATATTTCTTCATACCTGCATATATATATTTGGCGTCGCAAAATATGTATTTGTTGCACA	1138
OY		1260	CTATGTACAACAACACACTGAAGATATGTTTATAATAATTTGACTATTGGAAGTAAAA	1319
Db		1139	CTATGTACAACAACAACACTGAAGATATGTTTATAATAATTTGACTATTGGAAGTAAAA	1198
OY		1320	AAAAAAAAAAAA 1329	
Db		1199	AAAAAAAAAAAA 1208	
RESULT 5				
MMP1				
LOCUS			1428 bp	mRNA
DEFINITION				linear
ACCESION				ROD 12-SEP-1993
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
CDs				

```

misc_feature
polyA_site
/note="pot. polyA signal"
1401..1406
/note="polyA site"
1428
BASE COUNT      410 a      263 c      324 g      431 t
ORIGIN
Query Match      72.1%; Score 1013.2; DB 10; Length 1428;
Best Local Similarity 88.0%; Pred. No. 4,4e-156;
Matches 1204; Conservative 0; Mismatches 38; Indels 106; Gaps 5;

/db_xref="GI:55457"
/db_xref="MGD:MGI:97842"
/db_xref="SWISS-PROT:P11476"
/translation="WSSNRPDYDLFKLLILGDSGVKSCILLRFPADTYEISTT
GVDKIRITELDKGTIKLQIWDYDGOEFRTITSSYKNGAGIILVYDVTOESENK
KOMUEIDRYASENVNKLIVGNKCDLTKKVVDYTTAAEFDSIGIPLFESAKNAT
VEOSFPMMAAEIKRMRMGATAGAGSENVKIQSPVVKSGGGC"

58 GACGTTTGCTCTCCCGGACAGACCCATTCATTCCTTCTTCGATTAACCCGTGGCGGG 117
3 GACGTTTGCTCTACCGAAGACGCTTAGCTCATTCCTCCCTTCATTAAGTGGCGGG 62
118 AGAGTCAGGGGGGGGGCTGGCGGACAG-----GGCGGGGTGGCGGGCGGACGTG 172
63 AGAGTGGGGGGGGGGCTGGCGTCACAAAGGGGGTGGCGGGCGGGCGGCGACGTG 122
173 CAGTGACATGTCACAGCATGAATCCCGAATATGATATTAATCAAGTACTTGTATGG 232
123 CAGTGACATGTCACAGCATGAATCCCGAATATGATATTAATCAAGTACTTGTATGG 182
233 CGACTCAGGGGGTGGAAAGCTTGGCTTCTTCTTAAGTTGCGACATGATACATACGA 292
183 CGATTCTTGGGGTGGAAAGTCCCTGCTTCTCTTGGTTTCAGATGATAGTATACGA 242
293 AAGCTACATCGACACAATTTGGTGGATTTCAAAATTAAGACTATAGATTAGACGGGA 352
243 AAGCTACATCGACACAATTTGGTGGATTTCAAGTACGAATAGATTAGATGGAA 302
353 AACCAATCAAGCTTCAATA----- 371
303 AACCAATCAAGCTACAGATATGGGACACAGACGGCGGAAAGATTTCGAAATCACTTC 362
372 -----GATC 376
363 CAGTTATTACAGAGAGCCCATGGCATTATGTTGTATGATGTGACAGATCAGAGATC 422
377 CTTCATATATGTTAAACAGTGGCTGACAGAAATAGATGTTATGCCATGAAATGTCAA 436
423 CTTCATATACGTTAAACAGTGGCTGACAGAGATGATGCTACGCCAATGAAATGTCAA 482
437 CAAATGTTGGTAGGGAACAAATGTGATCTGACCAAGAAAGTAGTAGCTACACAA 496
483 CAAATGTTGGTAGGGAACAAATGTGATCTGACCAAGAAAGTAGTAGCTACACAA 542
497 AGCGAAGCAATTTGGTGTATCCCTTTGGAAATTCAGTTTGGAAACAGATGCTAAGATGC 556
543 AGCAAGCAATTTGGAGATTCCTTGGAAATTCAGTTTGGAAACAGATGCTAAGATGC 602
557 AACCAATTAACAGTCTTTCATGACATGGCAGCTAGATTAATAAAGCAATGGGTCC 616
603 AACCAATTAACAGTCTTTCATGACATGGCAGCTAGATTAATAAAGCAATGGGTCC 662
617 CGGAGCAACACTGGTGGTGGTGGAGAAATTCAGATTTAAATTTAGAGCACTCAGTCAA 676
663 TGGAGCTACACTGGTGGTGGTGGAGAAATTCAGATTTAAATTTAGAGCACTCAGTCAA 722
677 GCAGTCAGGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 736
723 GCAGTCAGGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 782
737 TGCATCTGAACCCCAAGTGAAGAAACAAATTTGCTGATTTGATCTATGTAGTGCAC 796
783 CGCATCTGAACCCCAAGTGAAGAAACAAATTTGCTGATTTGATCTATGTAGTGCAC 842

```

OY		797	TACAACAAATTCCTTACCAGCTGCTCCACAAGAAGTGCAAGATTGGTAATGGCAATACTGCAC-T	855
Dd		843	TACAACAAATTCCTTACCAGCTTTCCAC -AAGGTGAGAGATTGGTAATGGTCAATACTGCACCTT	901
OY		866	TTTTTTTTTTATTTCCCTTGAGCTCACAACAGCATCTTCATTTTCCGAACAGTGTTTAAACCTT	915
Dd		902	TTTTTTTTTTATTCCTTGAGCTCACAACCGGTAACTTCATTTTCCGAACAGTGTTTAAACCTT	961
OY		916	TGAGTGTGCGGTTTAAAAATATATGGTGTAATCTGTGTGCTTCCGTATACAGCACTGT	975
Dd		962	TGAGTGTGCGGTTTAAAAATATATGGTGTAATCTGTGTGCTTCCGTATACAGCACTGT	1021
OY		976	TTCCCGTGTGGTGGTAGAATATATTTGGTTTGAATCTTAAATGGCAAGTTTATGAGATGTC	1035
Dd		1022	TTCCCGTGTGGTGGTAGAATATATTTGGTTTGAATCTTAAATGGCAAGTGTTTATGAGATGTT	1081
OY		1036	AGGTTTATCTCTTCGAAGAATGAAGTACGCCATTTTGTATCAACACGCAACAGCATGTTC	1095
Dd		1082	GGGTTTATGCTCTTCGAAGAATGAAGTACGCCATTTTGTATCAACACGCAACAGCATGTTC	1141
OY		1096	TGTGCACTTTCATCATATAAGTTTAGTGAAGATGTATATGTAAGATCTGATTTGCTGATGT	1155
Dd		1142	TGTGCACTTTCATCATATAAGTTTAGTGAAGATGTATATGTAAGATCTGATTTGCTGATGT	1201
OY		1156	CCTTCCTGTAGAGTTATATAATGAAAAGATTACACATCTATGATTAATAGTTTCTTCACTCT	1215
Dd		1202	CCTTCCTGTAGAGTTATATAATGAAAAGATTACACATCTATGATTAATAGTTTCTTCACTCT	1261
OY		1216	CTGCATATATATTTGTGGCGCACGAATATTTGTATTTGTGCACACTATGTATACAAACAA	1275
Dd		1262	CTGCATATATATTTGTGGCGCACGAATATTTGTATTTGTGTGCACACTATGTATACAA--NA	1318
OY		1276	CTGAGAGATATGTTTAAATATATTTGTACTATTGTGGAAGTAAAAAANA	1323
Dd		1319	CTGAGAGATATGTTTAAATATATTTGTACTATTGTGGAAGTAAAAATATATCTCAA	1366
RESULT 6				
AF226873			2686 bp mRNA linear ROD 13-FEB-2000	
DEFINITION			Mus musculus small GTP-binding protein RAB1A mRNA, complete cds.	
ACCESSION			AF226873	
VERSION			AF226873.1 GI:96969621	
KEYWORDS				
SOURCE			Mus musculus.	
ORGANISM			Mus musculus.	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.	
TITLE			1 (bases 1 to 2686)	
JOURNAL			Wu.G. and Dorn.G.W. II.	
REFERENCE			Mouse RAB1A, member of RAS oncogene family mRNA	
AUTHORS			Unpublished	
TITLE			2 (bases 1 to 2686)	
JOURNAL			Wu.G. and Dorn.G.W. II.	
REFERENCE			Direct Submision	
AUTHORS			Submitted (19-JUN-2000) Internal Medicine, University of	
TITLE			Cincinnati, 231 Bethesda Avenue, Cincinnati, OH 45267-0590, USA	
JOURNAL				
FEATURES			Location/Qualifiers	
source			1..2686	
			/organism="Mus musculus"	
			/db_xref="taxon:10090"	
			29..646	
			/note="member of Ras oncogene family"	
			/codon_start=1	
			/product="small GTP-binding protein RAB1A"	
			/protein_id="AAP3844.1"	
			/db_xref="GI:6968622"	
			/translation="MSKAPPEPDYFLKLLIGDSGVGSCLLRFADDTYESITSTT	
			GVMDIKEDLKQTLQIMDPAQEKERFTITSYSIRRAHGLIVVDIDESFNV	
			KVMLEIDRASENNKLIVGNRCUETTKRVYTFAEPFSDISGLIEPLETSAKNATN	
			VEOSFEPMAEILRKMGEGATFGAGEKSNSVNIISTPVKSGGGCC"	
BASE COUNT			778 a 520 c 563 g 825 t	

ORIGIN	Query Match	68.0%	Score 955	DB 10	Length 2686
	Best Local Similarity	88.5%	Pred. No 1,1e-144		
	Matches 1123	Conservative 0	Mismatches 43	Indels 101	Gaps 4
Oy	152	CGGTGGCGGGCGGGCGGACCTGACATGTCAGACATGTCACGATGAAATCCCGAATATATATTT	211		
Db	1	CGGGCGGGCGGGCGGACGCTGCGAGTGCAGATGTCACGATGCAATGCCAATATGATATTTT	60		
Oy	212	ATTCAAGTTACTTCTGATTTGGCGACTAGGAGGGTTGAAAAGTTCCTTCCTTTAGGTT	271		
Db	61	ATTCAAGTTACTTCTGATTTGGCGACTTCTGGGGTTGGAAAAGCTTCCTTCCTTTAGGTT	120		
Oy	272	TGCAGATGATCATATPACGAAAGCTACATACACATTTGATGTGATTTCAAAATAG	331		
Db	121	TGCGAGTGTATCGTATACGAAAGCTCATATACACCAATTTGATGTGATTTCAAGATAG	180		
Oy	332	AACATAGAGTTAGACGGGAAAACATCAACAGCTTCAATA	371		
Db	181	AACATAGAGTTAGATGGGAAAACATCAACAGCTACAGATATGGGACACAGAGCCAGGA	240		
Oy	372	-----	371		
Db	241	AAGATTTGCAATCACTTCCTCCAGTTTATACAGAGAGCCCATGGCATATATGTGTGA	300		
Oy	372	-----GAGTCTTCAATATGTATTAACAGTGCCTGCGAAATATAGTGTG	415		
Db	301	TGATGTGACATCAGAGAGTCTTCAATATACGTATTAACAGTGCCTGCGAGGATATAGTGTG	360		
Oy	416	TTATNGCAGTAAATATGTCACAAATTTGTTGGTAGGAAACAAATGTATGTACACACAA	475		
Db	361	CTACGCGCATGAAATATGTCACAAAGTTGTTGGTAGGAAACAAATGTATGTACACACAA	420		
Oy	476	GAAGATGATACATACACACAGGAGAAATTCGTATTCCTTGGAAATTCGGTTT	535		
Db	421	GAAGATGATACATACACACAGGAGAAATTCGTATTCCTTGGAAATTCATTTT	480		
Oy	536	GGAACCATGCTTAAGATATGCAAGATATGAACAAGCTTTCAATGATGATGCACTGA	595		
Db	481	GGAACCATGCTTAAGATATGCAAGATATGAACAAGCTTTCAATGATGATGCACTGA	540		
Oy	596	GATTTAAAAAGGAATGGGTCCCGGAGACACAGCTGTGTGTGTGACAAATGTTAA	655		
Db	541	GATTTAAAAAGGAATGGGTCCCGGAGATACAGCTGTGTGTGTGACAAATGTTAA	600		
Oy	656	AATTCAGACACTCCATCAGATCAGATCAGCTGAGAGTTGCTTAAATTTGGCTCCATC	715		
Db	601	AATTCAGACACTCCATCAGATCAGATCAGCTGAGAGCTGCTTAAATTTGGCTCCATC	660		
Oy	716	CTTTTTCACAGCATATGAAATTTGCAACTCAACCAATGAAAAACAAATTTCCCTGAA	775		
Db	661	CTTTTTCACAGCATATGAAATTTGCAACTCAACCAATGAAAAACAAATTTCCCTGAA	720		
Oy	776	TTGTACTGATGTAGCTGCAGTACACACAGATTTCCAGCTGTCCACAAAGGTCCAGAGAT	835		
Db	721	TTGTACTGATGTAGCTGCAGTACACACAGATTTCCAGCTGTCCAC -AAGGTCCAGAGAT	779		
Oy	836	GTAATAGGCAATACAGAC -TTTTTTTTTATCCCTGACACGAAAGCAATACAT	894		
Db	780	GTAATAGGCAATACAGACTTTTTTTTTTATCCCTGACACGAAAGCAATACAT	839		
Oy	895	TTGAGAAAGCTTTAAACCTTTGTGTGTGTGTATATAATATGTGTATCTCTGT	954		
Db	840	TTGAGAAAGCTTTAAACCTTTGTGTGTGTGTATATAATATGTGTATCTCTGT	899		
Oy	955	GCTTTCCTGATACAGACAGCTGTTCCCGTGTGTGTGACAAATATTTTGTGTGTGTGT	1014		
Db	900	GCTTTCCTGATACAGAGTGTTCCTCCGCTGTGTGTGTGATAATATTTTGTGTGTGTGT	959		
Oy	1015	ATATGTGCAATTTAGATGTACAGGTTTAGTCTCTGGAAGTGAAGTTCAGCATTTTGTGA	1074		
Db	960	ATATGTGCAATTTAGATGTGTGGGTTTAGTCTCTGGAAGTGAAGTTCAGCATTTTGTGA	1019		





```

QY      863 TATTCCTGACACAGACAGCTACCTCATTTTCAGAACTGTTTAAACCTTGTGTGC 922
Db      76881 TATTCCTGACACAGACAGCTACCTCATTTTCAGAACTGTTTAAACCTTGTGTGC 76940
QY      923 TGGTTTAAATATATGTCATCTGTTGCTTCCGATATCCAGACTGTTCCCGT 982
Db      76941 TGGTTTAAATATATGTCATCTGTTGCTTCCGATATCCAGACTGTTCCCGT 77000
QY      983 GGTGGTTAGAAATATATTTGTTTGAATTTATATGTCATGTT----- 1027
Db      77001 GGTGGTTAGAAATATATTTGTTTGAATTTATATGTCATGTTGGGTTTGA 77060
QY      1028 ----- 1027
Db      77061 GTTGGGATTTTACCTGACGCTGACGCGCTTCCACAGCCAGGCCCTGGGTTGCG 77120
QY      1028 -----TAGATGTCAGGTTTACGCTTCTGAAGATGAAGTTCAAGCC 1066
Db      77121 TCCCGACCTCGAAAAAATAATAGATGTTGGGTTTACGCTTCTGAAGATGAAGTTCAAGCC 77180
QY      1067 ATTTTGTATCAACACAGCAGCAGCTGCTGCTCACTTCCATGCATGAAGTTTACTGAGA 1126
Db      77181 ATTTTGTATCAACAAATAGCAGCAGCTGCTGCTCACTTCCATGCATGAAGTTTACTGAGA 77240
QY      1127 TGTATATGTAAGATGCTGATTTGCTAGTTCTTCTGAGAGTTAATGAAGATTA 1186
Db      77241 CATATATGTAAGATGCTGATTTGCTAGTTCTTCTGAGAGTTAATGAAGATTA 77299
QY      1187 CACATATGATTAATAGTTTCTTCTGATCTGCTGATTAATTTGGCGCAGATTAATGT 1246
Db      77300 CACATATGATTAATAGTTTCTTCTGATCTGCTGATTAATTTGGCGCAGATTAATGT 77359
QY      1247 AATTTGTGACACACTATGTAACAAACAGCTGAAGATATGTTAATAATTTACTACTA 1306
Db      77360 AATTTGTGACACACTATGTAACAA--AACTGAAGACATGTTAATAATTTACTACTA 77416
QY      1307 TTGGAAGTAAAAAAA 1323
Db      77417 TCGGAAGTAAATATCAAA 77433

RESULT 8
AC007318/ 204230 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-420C9 from 2, complete sequence.
DEFINITION AC007318
ACCESSION AC007318.4 GI:13435282
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 204230)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 204230)
AUTHORS Wollan, A., Hawkins, M., Kallick, J., Harfins, R. and Lehnert, L.
TITLE The sequence of Homo sapiens BAC clone RP11-420C9
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 204230)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 204230)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

```

REFERENCE MO 63108, USA
AUTHORS 5 (bases 1 to 204230)
TITLE Waterston, R.
JOURNAL Direct Submission
SUBMITTED (09-AUG-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 204230)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 23, 2001 this sequence version replaced g1:7630984.
COMMENT ----- Genome Center.
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
Summary Statistics
Center project name: H_NH0420C09

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frangen, E.,
Tateno, M., Catalano, J.J., and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-350H15; the clone sequenced
to the right is AC034121. Actual start of this clone is at base
position 1 of RP11-420C9; actual end is at base position 204230 of
RP11-420C9.

The sequence RP11-420C9 from base position 44230 to 46506 consists
of PCR product from clone DNA. Base position 44700 to 44774
represents sequence of a single plasmid subclone derived from PCR
amplification of clone DNA.
FEATURES
Location/Qualifiers
1..204230
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-420C9"
/clone_id="RPCI-11"
179..486
/feat_family="Alu"
890..1083
/feat_family="MER1-type"
1458..1789
repeat_region
repeat_region
repeat_region

```



QY 1162 TGTAGCTTAAATGGAAGATACATATCTGATTAAATAGTTCTTCATCTGTCAT 1221  
|||||  
Db 9676 TGTAGCTTAAATGGAAGATACATATCTGATTAAATAGTTCTTCATCTGTCAT 9617  
QY 1222 ATATTTGTGCTGCAGAAATTTTGTATTTTGTGCACACTATGTAACAAACACTGAG 1281  
|||||  
Db 9616 ATATTTGTGCTGCAGAAATTTTGTATTTTGTGCACACTATGTAACAAACACTGAG 9557  
QY 1282 ATATTTGTATTAATATTTGTTACTATTTGGAAGTAAAAAAA 1323  
|||||  
Db 9556 ATATTTGTATTAATATTTGTTACTATTTGGAAGTAAAAAAA 9515

RESULT 9  
AL606522 203071 bp DNA linear ROD 22-MAY-2002  
LOCUS Mouse DNA sequence from clone Rp23-19L22 on chromosome 11, complete  
DEFINITION sequence.  
ACCESSION AL606522 GI:20792609  
VERSION AL606522.6 GI:20792609  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 203071)  
Almeida, J.  
AUTHORS  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (22-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
humquerry@sanger.ac.uk  
On May 14, 2002 this sequence version replaced gi:11968263.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP  
database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Rp23-19L22 is  
from the Rp23 Mouse PAC Library  
constructed by the group of Pieter de Jong.  
For further details see <http://www.chori.org/dacpac/home.htm>  
VECTOR: pBAC3.6.

FEATURES  
source location/Qualifiers  
1..203071  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="Rp23-19L22"  
/clone\_1lb="Rp23-19L22"  
BASE COUNT 57371 a 44818 c 43421 g 57461 t  
ORIGIN

Query Match 52.8%; Score 741.4; DB 10; Length 203071;  
Best Local Similarity 96.2%; Pred. No. 3.5e-110;  
Matches 792; Conservative 0; Mismatches 26; Indels 5; Gaps 3;

QY 502 AGGAATTTGCTGATTCCTTGGATTCGGTTTGGAAACAGAGTCTAAGATGCAAGA 561  
|||||  
Db 177379 AGGAATTTGCTGATTCCTTGGATTCGGTTTGGAAACAGAGTCTAAGATGCAAGA 177438  
|||||  
QY 562 ATGTAGAACAGTCTTTATGACGATGCGACGTGAGATTAACGCAATGGGTCGCGAG 621

Db 177439 ATGTAGAACAGTCTTTATGACGATGCGACGTGAGATTAACGCAATGGGTCGCGAG 177498  
|||||  
QY 622 CACAGAGCTGGTGGCTAGAGATTCATGTTAAATTCACAGCACTGCATCAAGCAGT 681  
|||||  
Db 177499 CTACAGCTGGTGGCTAGAGATTCATGTTAAATTCACAGCACTGCATCAAGCAGT 177558  
|||||  
QY 682 CAGGTGAGAGGTGGCTAGAGATTCATGTTAAATTCACAGCACTGCATCAAGCAGT 741  
|||||  
Db 177559 CAGGTGAGAGGTGGCTAGAGATTCATGTTAAATTCACAGCACTGCATCAAGCAGT 177618  
|||||  
QY 742 TCTGAAACCAAGTGAAGAAAAACAAATTCCTGAATGTTACTGTATGAGTGCACACTACAA 801  
|||||  
Db 177619 TCTGAAACCAAGTGAAGAAAAACAAATTCCTGAATGTTACTGTATGAGTGCACACTACAA 177678  
|||||  
QY 802 CAGATTCCTTACCGTCTCCACAAAGGTGAGATGTAATGTCATCTGAC--TTT 860  
|||||  
Db 177679 CAGATTCCTTACCGTCTCCAC--AAGGTGAGATGTAATGTCATCTGAC--TTT 177737  
|||||  
QY 861 TTTATTCCTTACCGTCTCCACAGACGCTAATCTGATTTTCAAGCTGTTTAACTTTGCT 920  
|||||  
Db 177738 TTTATTCCTTACCGTCTCCACAGACGCTAATCTGATTTTCAAGCTGTTTAACTTTGCT 177797  
|||||  
QY 921 GCTGGTTTAAATATATGTTGTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 980  
|||||  
Db 177798 GCTGGTTTAAATATATGTTGTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 177857  
|||||  
QY 981 GTGGTGGTTGAATATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1040  
|||||  
Db 177858 GTGGTGGTTGAATATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 177917  
|||||  
QY 1041 TAGCTCTTGAAGATGAAGTTCAGCCATTTTGTATGTAACAAACAGCAAGCACTGCTGTA 1100  
|||||  
Db 177918 TAGCTCTTGAAGATGAAGTTCAGCCATTTTGTATGTAACAAACAGCAAGCACTGCTGTA 177977  
|||||  
QY 1101 CTTTCATGCTAATAGTTTATGAGATTTATGTAAGATCTGATTTGCTAGTTCTCC 1160  
|||||  
Db 177978 GTTTCACGCAATGAAGTTCAGCTTTATGTAAGATCTGATTTGCTAGTTCTCC 178037  
|||||  
QY 1161 TTGATAGATTTAAATGGAAGATTTACATCTGATTAATGTTCTTCATCTACCTGCA 1220  
|||||  
Db 178038 TGTAGATTTAAATGGAAGATTTACATCTGATTAATGTTCTTCATCTACCTGCA 178097  
|||||  
QY 1221 TATAATTTGCTGTCAGAAATATTTGTAATTTGTTGTCACACTATGTAACAAACACTGCA 1280  
|||||  
Db 178098 TATAATTTGCTGTCAGAAATATTTGTAATTTGTTGTCACACTATGTAACAA--AATCGAA 178154  
|||||  
QY 1281 GATATGTTTAAATATTTGTTACTTATTTGGAAGTAAAAAAA 1323  
|||||  
Db 178155 GATATGTTTAAATATTTGTTACTTATTTGGAAGTAAAAATAATATCAAA 178197  
|||||

RESULT 10  
AL512349 116614 bp DNA linear HTG 10-JUL-2001  
LOCUS Homo sapiens chromosome 6 clone RP11-223J2, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AL512349  
VERSION AL512349.7 GI:13121473  
KEYWORDS HTG; HTGS\_PHAZE2; HTGS\_CANCELLED.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 116614)  
Bates, K.  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk) clone  
requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Feb 23, 2001 this sequence version replaced gi:12556581.  
COMMENT  
----- Genome Center  
Center: Sanger Centre



```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humbrey@sanger.ac.uk
-----
Project Information
Center project name: ba22332
-----
Summary Statistics
Assembly program: xgtdp, version 4.5
Sequencing vector: plasmid, 108752; 100% of reads
Chemistry: dye-terminator, big dye, 100% of reads
Consensus quality: 11589 bases at least Q40
Consensus quality: 11635 bases at least Q30
Consensus quality: 11647 bases at least Q20
Insert size: 11634; sum of contigs
Insert size: 11670; 7.2% error; agarose-tp
Quality coverage: 6.23x in Q20 bases; sum-of-contigs quality
coverage: 6.37x in Q20 bases; agarose-tp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
    source
        1. 116614
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone="RP11-22332"
            /clone_11b="RPC1-11.1"
            1. 116614
            /note="assembly_fragment:02232
            clone_end:SP6
            vector_slide:right"
misc_feature
    misc_count
        34471 a 21866 c 22518 g 37753 t 6 others
ORIGIN
Query Match
    48.9%: Score 687.6; DB 2; Length 116614;
    Best Local Similarity 74.9%; Pred. No. 1.6e-101;
    Matches 1099; Conservative 0; Mismatches 219; Indels 150; Gaps 13;
    Db
        27 TGATTTGCTTCRAGGGAGAGAGTACGGGAACACTTTCCTCCCGGAACGCTTATCT 86
        10557 TGATTTGCTTCRAGGGAGAGAGTACGGGAACACTTTCCTCCCGGAACGCTTATCT 10498
    Oy
        87 CATTCCTTCTTTCATCATCCCTGGGCGGGAGACTCAGGGCGGGCTGGCGGACGAC 146
        10497 CATTCCTTCTTTCATCATCCCTGGGCGGGAGACTCAGGGCGGGCTGGCGGACGAC 10467
    Db
        147 GGGCGGGCTGGCGGGCGGGAGCTGCATGCATGCATGCATGCATGCATGCATGCATGCAT 206
        10466 AGTCATGCTGGCGGGAGAGAGCTGCATGCATGCATGCATGCATGCATGCATGCATGCAT 10407
    Oy
        207 TATTATTACAGTTCCTCTGATTTGGCGAGCTAGGGGCTGGGAAGTCTGGCTCTTCTT 266
        10406 TATTATTACAGTTCCTCTGATTTGGCGAGCTAGGGGCTGGGAAGTCTGGCTCTTCTT 10347
    Db
        267 AGGTTTCAGATG-----ATACATATACAGAAAGTACATCAGACACATTTGGTGTG 317
        10346 AGGTTTCAGATGAAACATGTAAATACATATATACAGAAAGTACATCAGACCAATTAAGTGG 10287
    Oy
        318 GATTTCAAAATAGAACATATAGAGTTAGACGGGAACCAATCAAGCTTCATAATA----- 371
        10286 AATTTCAAAATAGAACATATATATATATATATATATATATATATATATATATATATATAT 10227
    Oy
        372 ----- 371
    Db
        10226 ACAACAGCCGAGGAAGATTTTATCATCACTCAGTTATATACAGAGAGCCCAAGGCTGTA 10167
    Oy
        372 -----GAGTCCTCAATAATTTTAACTAGTGTGCTCAGAGAAATATGATG 415
    Db
        10166 TGATATACAGATCAGAGAGTCTTCATATATTTTAAATAGTGGTTCAGAGAAATATGATCA 10107
    Oy
        416 TTATGCCAGTGAATATGCACAATTTGTTGGTAGGCAACAATTGATGTGACCAACA 475
    ||

```

[illegible]

ACCESSION AL049545  
 VERSION AL049545.6 GI:5002650  
 KEYWORDS HTG; 60S Ribosomal Protein L7; RAB1; RAS; RPL7.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 104081)  
 AUTHORS Smith, S.  
 JOURNAL Direct Submission  
 Submitted (06-JUL-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 Requests: clonerequests@sanger.ac.uk  
 On Jun 7, 1999 this sequence version replaced gi:4835284.  
 During sequence assembly data is compared from overlapping clones. The following differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the validation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
 Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL  
 This sequence is the entire insert of clone 263J7. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
 This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>  
 263J7 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCIPAC2>.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..104081  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="q14.3-15"  
 /clone="RPI-263J7"  
 /clone\_id="RPCI-1"  
 1..4880  
 /note="LIP5 repeat: matches -237. .5623 of consensus"  
 4895..5820  
 /note="LIP5 repeat: matches -1405. .-405 of consensus"  
 5861..6011  
 /note="WTR17 repeat: matches 36. .174 of consensus"  
 6032..6086  
 /note="L1 repeat: matches 3626. .3680 of consensus"  
 6740..8505  
 /note="LIP12 repeat: matches 4282. .6165 of consensus"  
 8512..9328  
 /note="LIP repeat: matches 3049. .3866 of consensus"  
 9329..9626  
 /note="AluYb8 repeat: matches 1. .317 of consensus"  
 9627..10966  
 /note="LIP repeat: matches 1712. .3049 of consensus"  
 complement(11030..11749)  
 /gene="dJ263J7.1"  
 /gene="dJ263J7.1"  
 pseudogene); match: CDNAS: Em:X52967 Em:X57959 Em:U16558  
 Em:X57958 Em:M85235 Em:M17422 Em:M29016 Em:X57961  
 Em:X57960 Em:U63785; match: ESTs: Em:AL047212 Em:AI683841  
 Em:AI609411 Em:AI620251 Em:AI678713 Em:AA853290  
 Em:AI188448 Em:AI246632 Em:E29697.1 Em:AA640941  
 Em:AA147085 Em:AI354542 Em:AA68558 Em:AA172041

Em:AA736700 Em:AA181027 Em:AA617963 Em:AA129663  
 Em:AI385514 Em:AA586843 Em:AA148321 Em:AA147910  
 Em:AA065193 Em:AA223084 Em:AA065240 Em:AA314610  
 Em:AI498784 Em:AA313779 Em:AA626741 Em:AA397390  
 Em:AA639843 Em:AA598466 Em:AA567875 Em:AA314579  
 Em:AA081412 Em:AA638601 Em:AA129729 Em:AA236303  
 Em:AA580325 Em:AA147597 Em:AA313220 Em:AA516160  
 Em:AA311117 Em:AA562608 Em:AA635115 Em:AI367854  
 Em:AA308128 Em:AA846920 Em:AA608110 Em:AA745100  
 Em:AA607384 Em:AA507813 Em:AA314838 Em:AA736534  
 Em:AA591026 Em:AA008632 Em:W77465 Em:AA648073 Em:AA648202  
 Em:AA652160 Em:AA715370 Em:D51897 Em:D51979 Em:D52053  
 Em:AA760175 Em:AA856854 Em:N98334 Em:AA649459 Em:AA879164  
 Em:AA332842 Em:AA886888 Em:AA173529 Em:AA091272  
 Em:AA033750; match: proteins: Sw:P18124 Sw:P14148  
 Sw:P05426 Sw:Q42208 Sw:Q01802 Sw:Q60143 Sw:P25457  
 Sw:U18674 Sw:F32100 Sw:F05737 Sw:Q12213 Tr:Q14371  
 Tr:Q08953"  
 /pseudo  
 /codon\_start=1  
 /evidence-not\_experimental  
 11307..11789  
 /note="match: GSS: AQ081091 B54651 AQ028779"  
 11790..12860  
 /note="LIP13 repeat: matches 647. .1730 of consensus"  
 12908..13190  
 /note="AluY repeat: matches 3. .302 of consensus"  
 13341..13560  
 /note="HAI1 repeat: matches 1483. .1721 of consensus"  
 13346..13347  
 /note="3 subclones read TCTT, 2 read TCT"  
 13594..15880  
 /note="LIP2 repeat: matches 3950. .6154 of consensus"  
 15881..16608  
 /note="LIP4 repeat: matches 5413. .6144 of consensus"  
 16609..19090  
 /note="LIP2 repeat: matches 1394. .3950 of consensus"  
 19096..19323  
 /note="LIP3 repeat: matches 3. .231 of consensus"  
 19472..23244  
 /note="LIP repeat: matches 638. .4389 of consensus"  
 23240..23462  
 /note="LIPa repeat: matches -1549. .-1330 of consensus"  
 23577..29885  
 /note="LIP3 repeat: matches -218. .6130 of consensus"  
 29886..29943  
 /note="LIPa5 repeat: matches 6081. .6139 of consensus"  
 29944..30234  
 /note="LIP3 repeat: matches -508. .-218 of consensus"  
 complement(30558..30912)  
 /note="match: EST AA528504"  
 31385..31601  
 /note="MIR repeat: matches 29. .256 of consensus"  
 31721..32002  
 /note="LIP7 repeat: matches 5862. .6145 of consensus"  
 32279..32583  
 /note="AluX repeat: matches 1. .308 of consensus"  
 33356..33592  
 /note="MIR repeat: matches 8. .247 of consensus"  
 33779..37507  
 /note="LIM2 repeat: matches 2533. .6308 of consensus"  
 37508..37811  
 /note="AluY repeat: matches 3. .305 of consensus"  
 37812..37985  
 /note="LIM2 repeat: matches 2362. .2533 of consensus"  
 37976..41630  
 /note="LIM repeat: matches -1390. .2698 of consensus"  
 41878..41924  
 /note="MERB repeat: matches 127. .173 of consensus"  
 42325..42418  
 /note="MAD1 repeat: matches 1. .79 of consensus"  
 42451..42853  
 /note="WTR1C repeat: matches 23. .466 of consensus"

[illegible]

```

OY      1307 TTGCAAGTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1366
DB      94457 CAATGGATCTAATTAATTAACCAAGAGCTTCGTGCAGAAAAATAACAGCAGAGAGTG 94398
OY      1367 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1394
DB      94397 AATAGACACACACCAAAATGAGAGAAA 94370

RESULT 12
AC067945/c 162471 bp DNA linear PRI 07-NOV-2001
LOCUS      AC067945
DEFINITION Homo sapiens BAC clone RP11-629B4 from 2, complete sequence.
ACCESSION  AC067945
VERSION     AC067945.4 GI:13786486
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 162471)
            Sulston, J.E. and Waterston, R.
            Toward a complete human genome sequence.
            Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL    99063792
PUBMED     9847074
REFERENCE  2 (bases 1 to 162471)
            Isak, A., Elliott, G., Doebber, A., Abbott, A., Hawkins, M. and Falk, A.
            The sequence of Homo sapiens BAC clone RP11-629B4
            Unpublished
JOURNAL    3 (bases 1 to 162471)
            Waterston, R.H.
            Direct Submission
            Submitted (27-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            4 (bases 1 to 162471)
            Waterston, R.H.
            Direct Submission
            Submitted (25-APR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            5 (bases 1 to 162471)
            Waterston, R.
            Direct Submission
            Submitted (09-AUG-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            6 (bases 1 to 162471)
            Waterston, R.
            Direct Submission
            Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Apr 25, 2001 this sequence version replaced g1:13518285.

COMMENT
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
-----
Summary Statistics
Center project name: H_NH0629B04

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanesse, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

#### VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-45219, 200 bp overlap; the clone sequenced to the right is RP11-31739. Actual start of this clone is at base position 195 of RP11-629B4; actual end is at base position 162471 of RP11-629B4.

#### FEATURES

##### source

```

1..162471
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-629B4"
/clone_1lb="RP11-11"
/rpt_family="AT_rich"
777..1067
/rpt_family="Alu"
1352..1380
/rpt_family="AT_rich"
1887..2071
/rpt_family="MIR"
3209..3257
/rpt_family="MIR"
4158..4475
/rpt_family="L1"
4476..4775
/rpt_family="Alu"
4776..5044
/rpt_family="L1"
5108..5242
/rpt_family="Alu"
5428..5465
/rpt_family="AT_rich"
5617..5910
/rpt_family="Alu"
5997..6133
/rpt_family="Alu"
6135..6178
/rpt_family="Alu"
6213..6498
/rpt_family="Alu"
9226..9268
/rpt_family="AT_rich"
9573..9600
/rpt_family="AT_rich"
10528..10562
/rpt_family="AT_rich"
11440..11467
/rpt_family="AT_rich"
12297..12325
/rpt_family="AT_rich"
12472..12518
/rpt_family="TTATATG)n"
14656..14747
/rpt_family="CRI"
16030..16089

```

repeat_region	/rpt_family="AT-rich"	17221..17290	DB	69889	AAGCAATAGCTGAGTGGCAGCGGCTGCTGATGCTGCTCCAGGGGCGAGAGTCGGGGAGA	69830
repeat_region	/rpt_family="L2"	17828..17863	QY	60	CGTTGCTCTCCGGGAACGCTATCTATCTCTTCTTTCGATACCGCGCGGAG	119
repeat_region	/rpt_family="L2"	17964..18138	DB	69829	CATTACTCTCCAGGAACGCTATCTATCTCTTCTTTCGATACCGCGCGGAG	69770
repeat_region	/rpt_family="MER1-type"	18139..18209	QY	120	AGTCAGGGCGGGCGCTCGGAGCAAGGGCGGGGTGGCGGGCGGAGCTGAGTAC	179
repeat_region	/rpt_family="L2"	18210..18521	DB	69769	AGTCAGGGCGGAGCTGAGTACGAGCAAGGGGTGGCGGGCGGAGCTGAGTAC	69710
repeat_region	/rpt_family="Alu"	18522..18865	QY	180	ATGTCACAGTGAATCCGCAATATGATATATATCAAGTACTCTGATGGCGACTA	239
repeat_region	/rpt_family="L2"	19330..19710	DB	69709	ATGTCACAGTGAATCCGCAATATGATATATATCAAGTACTCTGATGGCGACTA	69651
repeat_region	/rpt_family="L2"	20271..20291	QY	240	GGGGTGGAAAGTCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	298
repeat_region	/rpt_family="AT-rich"	21212..21327	DB	69650	GGGGTGGAAAGTCTGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	69591
repeat_region	/rpt_family="MER1-type"	21523..21828	QY	299	CATCAGCAGCAATTTGGTGGATTTTCAAAATPAGACTATPAGAGTTAGCGGAAACAT	358
repeat_region	/rpt_family="Alu"	22806..22860	DB	69590	CATCAGCAGCAATTTGGTGGATTTTCAAAATPAGACTATPAGAGTTAGCGGAAACAT	69531
repeat_region	/rpt_family="CT-rich"	22994..23105	QY	359	CAGGCTGAATA-----	371
repeat_region	/rpt_family="MER1-type"	25554..25846	DB	69530	CAGGCTGAATA-----	69471
repeat_region	/rpt_family="Alu"	25927..26475	QY	372	-----GAGTCTTCAA	382
repeat_region	/rpt_family="L2"	26500..26664	DB	69470	TTTACAGAGACCAATGCGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	69411
repeat_region	/rpt_family="L2"	26754..26897	QY	383	TTTACAGAGACCAATGCGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	442
repeat_region	/rpt_family="MER1-type"	26898..27113	DB	69410	TTTACAGAGACCAATGCGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	69351
repeat_region	/rpt_family="L2"	27513..27652	QY	443	GTGTTAGGGAACAAATGATGATGATGATGATGATGATGATGATGATGATGATG	502
repeat_region	/rpt_family="L2"	28002..28057	DB	69350	GTGTTAGGGAACAAATGATGATGATGATGATGATGATGATGATGATGATGATG	69291
repeat_region	/rpt_family="MER1-type"	28065..28196	QY	503	GGAATTTCTGATTCCTTGGAAATTCCTTTTGGAAACAGTCT-----	548
repeat_region	/rpt_family="Alu"	28481..28541	DB	69290	GGAATTTCTGATTCCTTGGAAATTCCTTTTGGAAACAGTCT-----	69231
repeat_region	/rpt_family="MER103"	30093..30138	QY	549	-----	548
repeat_region	/rpt_family="T(n)"	33232..33532	DB	69230	GGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	69171
repeat_region	/rpt_family="Alu"	36502..36816	QY	549	-----	548
repeat_region	/rpt_family="Alu"	37354..37634	DB	69170	CACGTGAGTACAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	69111
repeat_region	/rpt_family="Alu"	41942..41974	QY	549	-----	548
repeat_region	/rpt_family="L2"	42371..42439	DB	69110	AAATATCAAAATTTAGCAGAGCATGTGGGGGCGCTGTATCCAGCTACTTGGAG	69051
repeat_region	/rpt_family="MER1-type"	42641..42718	QY	549	-----	548
repeat_region	/rpt_family="T(n)"	42719..42746	DB	69050	CTGAGGAGAGAAATTTGCTCGAACCCAGGTGGGAGGTGATGATGATGATGATG	68991
repeat_region	/rpt_family="CA)n"	43845..43966	QY	549	-----	548
repeat_region	/rpt_family="L2"	47469..47686	DB	68990	CATTGCTCTCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	68931
repeat_region	/rpt_family="MER1-type"	47727..47789	QY	552	AATGCAAGCAATTTAGTACAGAGTCTTTCATGAGAGAGAGAGAGAGAGAGAG	611
repeat_region	/rpt_family="L2"	47840..48153	DB	68930	AATGCAAGCAATTTAGTACAGAGTCTTTCATGAGAGAGAGAGAGAGAGAGAG	68871
repeat_region	/rpt_family="L2"	47840..48153	QY	612	GTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	671
repeat_region	/rpt_family="L2"	47840..48153	DB	68870	GTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	68811
repeat_region	/rpt_family="L2"	47840..48153	QY	672	GTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	731
repeat_region	/rpt_family="L2"	47840..48153	DB	68810	GTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	68751

Query Match 48.7% Score 684.4; DB 9; Length 162471;  
 Best Local Similarity 71.1% Freq. No. 5.4e-10;  
 Matches 1237; Conservative 0; Mismatches 76; Indels 427; Gaps 8;



Db	2459	CCTGCTAGAGTATTAAGGACACGACACCTATCTGATTAATTTCTTCACTCTG	2518	REFERENCE	Baylór Plaza, Houston, TX 77030, USA
Qy	1219	CATATATGTCGCGCGAGATATTTCTATTTGTCACACTTGTGACAAACACTG	1278	3 (bases 1 to 207945)	
Db	2519	CAATATATTTGTCGCGAGATATTTGTTGTCACACTTGTGACAAACAACTG	2574	Worley, K.C.	
Qy	1279	AGATATGTTTAAATTAATTTGTTACTTATTTGGAAGTAAAAAAA	1323	Direct Submission	
Db	2575	AGATATGTTTAAATTAATTTGTTACTTATTTGGAAGTAAAAA	2619	Submitted (17-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
RESULT 14	AC117841	207945 bp	DNA	linear	HG 17-JUL-2002
LOCUS	Rattus norvegicus clone CH230-360J6, *** SEQUENCING IN PROGRESS				
DEFINITION	*** 73 unordered pieces.				
ACCESSION	AC117841				
VERSION	AC117841.4	GI:21746140			
KEYWORDS	HTG; HTGS; PHASE1.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 207945)				
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Blum, K., Blankenburg, K., Bonini, D., Bouck, J., Bowe, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Burch, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C., Carron, T.P., Carter, M., Cavazos, S.R., Chacko, J., Chaves, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyne, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dedrich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douhaite, K.J., Draper, H., Dugan-Pocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Gabisi, R., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Hulys, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, A., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, N., Nickerson, E., Nwokwesi, S., Ogihara, M., Ogihara, N., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojka, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanil, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleczky, R., Woodcock, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.				
TITLE	Direct Submission				
JOURNAL	2 (bases 1 to 207945)				
REFERENCE	Worley, K.C.				
AUTHORS	Direct Submission				
TITLE	Submitted (11-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One				

REFERENCE	3 (bases 1 to 207945)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (17-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jul 14, 2002 this sequence version replaced gi:20258099.
	Center: Baylor College of Medicine
	Center code: BCM
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>
	Contact: <a href="mailto:hgsc-help@bcm.tmc.edu">hgsc-help@bcm.tmc.edu</a>
	Project Information
	Center project name: GRP
	Center clone name: CH230-360J6
	Summary Statistics
	Sequencing vector: pBlasmid;
	Chemistry: Dye-terminator Big Dye; 100% of reads
	Assembly program: Phrap; version 0.990329
	Consensus quality: 129522 bases at least 040
	Consensus quality: 134968 bases at least 030
	Consensus quality: 138682 bases at least 020
	*****
	NOTE: Estimated insert size may differ from sequence length
	(see <a href="http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html</a> )
	NOTE: This is a 'working draft' sequence. It currently
	consists of 73 contigs. The true order of the pieces
	is not known and their order in this sequence record is
	arbitrary. Gaps between the contigs are represented as
	runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence.
	* as soon as it is available and the accession number will
	be preserved.
	1 1046: contig of 1046 bp in length
	1 1047 1146: gap of unknown length
	* 1147 2189: contig of 1037 bp in length
	* 2184 2283: gap of unknown length
	* 2284 3421: contig of 1138 bp in length
	* 3422 3521: gap of unknown length
	* 3522 4686: contig of 1165 bp in length
	* 4687 4786: gap of unknown length
	* 4787 6158: contig of 1372 bp in length
	* 6159 6258: gap of unknown length
	* 6259 7327: contig of 1069 bp in length
	* 7328 7427: gap of unknown length
	* 7428 9082: contig of 1655 bp in length
	* 9083 9182: gap of unknown length
	* 9183 10231: contig of 1049 bp in length
	* 10232 10331: gap of unknown length
	* 10332 11663: contig of 1333 bp in length
	* 11664 11763: gap of unknown length
	* 11764 13476: contig of 1713 bp in length
	* 13477 13576: gap of unknown length
	* 13577 15050: contig of 1474 bp in length
	* 15051 15150: gap of unknown length
	* 15151 16522: contig of 1372 bp in length
	* 16523 16623: gap of unknown length
	* 16623 18019: contig of 1397 bp in length
	* 18020 18119: gap of unknown length
	* 18120 19874: contig of 1655 bp in length
	* 19875 19876: gap of unknown length
	* 19876 21203: contig of 1329 bp in length
	* 21204 21303: gap of unknown length
	* 21304 22591: contig of 1688 bp in length
	* 22592 23091: gap of unknown length
	* 23092 24721: contig of 1630 bp in length
	* 24722 24821: gap of unknown length
	* 24822 25833: contig of 1012 bp in length
	* 25834 25933: gap of unknown length
	* 25934 27568: contig of 1635 bp in length
	* 27569 29226: gap of unknown length
	* 29226: contig of 1558 bp in length

29327	29326	gap of unknown length
29327	30423	contig of 1098 bp in length
30425	30524	gap of unknown length
30525	32145	contig of 1621 bp in length
32146	32245	gap of unknown length
32246	33498	contig of 1253 bp in length
33499	33598	gap of unknown length
33599	34724	contig of 1126 bp in length
34725	34824	gap of unknown length
34825	36502	contig of 1678 bp in length
36503	36602	gap of unknown length
36603	37971	contig of 1369 bp in length
37972	38071	gap of unknown length
38072	39263	contig of 1192 bp in length
39264	39363	gap of unknown length
39364	41128	contig of 1765 bp in length
41129	41228	gap of unknown length
41229	42650	contig of 1422 bp in length
42651	42750	gap of unknown length
42751	44150	contig of 1400 bp in length
44151	44250	gap of unknown length
44251	45440	contig of 1190 bp in length
45441	45540	gap of unknown length
45541	47375	contig of 1835 bp in length
47376	47475	gap of unknown length
47476	48634	contig of 1159 bp in length
48635	48734	gap of unknown length
48735	50080	contig of 1346 bp in length
50081	50180	gap of unknown length
50181	51471	contig of 1291 bp in length
51472	51571	gap of unknown length
51572	52959	contig of 1388 bp in length
52960	53059	gap of unknown length
53060	54842	contig of 1783 bp in length
54843	54942	gap of unknown length
54943	57384	contig of 2442 bp in length
57385	57484	gap of unknown length
57485	59403	contig of 1919 bp in length
59404	59503	gap of unknown length
59504	60963	contig of 1460 bp in length
60964	61063	gap of unknown length
61064	62330	contig of 1267 bp in length
62331	62430	gap of unknown length
62431	63700	contig of 1270 bp in length
63701	63800	gap of unknown length
63801	65110	contig of 1310 bp in length
65111	66210	gap of unknown length
66211	66895	contig of 1685 bp in length
66896	66995	gap of unknown length
66996	68391	contig of 1396 bp in length
68392	68491	gap of unknown length
68492	70672	contig of 2181 bp in length
70673	70772	gap of unknown length
70773	72373	contig of 1601 bp in length
72374	72473	gap of unknown length
72474	73328	contig of 2855 bp in length
73329	75428	gap of unknown length
75429	77871	contig of 2443 bp in length
77872	77971	gap of unknown length
77972	80743	contig of 2772 bp in length
80744	80843	gap of unknown length
80844	82486	contig of 1643 bp in length
82487	82586	gap of unknown length
82587	84732	contig of 2146 bp in length
84733	84832	gap of unknown length
84833	87219	contig of 2387 bp in length
87220	87319	gap of unknown length
87320	90363	contig of 3044 bp in length
Query Match	48.3%: Score 679.2; DB 2: length 207945;	
Best local Similarity	81.9%: Pred. No. 3.7e-100;	
Matches 937; Conservative	0; Mismatches 98; Indels 109; Gaps 9;	

Qy	1	AAGGATAGCTGAGT-CCGGGCGCTGCTGATTTGTTGTTACAGGAGGAGTGGAGGA	59
Db	91854	AAGGATACCTGAGTGGTGGCGGCTGCTGATTTGTTAAGGCTGGCT-GGGCTTGA	91796
Qy	60	CGTTGGCTCTCCGGAACACCCATGCTCATTCCTTCTTTCATTCCTGCTGGCGGAG	119
Db	91795	CGTTGGCTCTACGGAACACCTTACCTTCCCTCCCTCCATTAACCTGTGACAGA	91736
Qy	120	AGTCAGGCGGCGGCTGGCGGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	179
Db	91735	GCTTGGGCGGCGGCGGCTGGCGGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	91679
Qy	180	ATGTCAGGATGGAATCCCGGAATGATTTATTTATGAACTTCTTGTGTTGGCACTGA	239
Db	91678	ATGTCAGGATGGAATCCCGGAATGATTTATTTATGAACTTCTTGTGTTGGCACTGA	91619
Qy	240	GGGTTTGGAAAGCTTGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	



DB 90783 GTCGTTGTTAGCAATATTTGTTGATGTTTATTTGTCAGTTTATGATTTGGCTT 90724  
 OY 1041 TAGT 1044  
 DB 90723 TAGT 90720

RESULT 15  
 AC099354/C  
 LOCUS  
 DEFINITION AC099354 174028 bp DNA linear HTG 11-JUL-2002  
 Rattus norvegicus clone CH230-901, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 67 unordereed pieces.  
 AC099354  
 AC099354.3 GI:21722818  
 HTG: HTGS-PHASE1.  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 174028)  
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbata,J., Benton,J., Binsage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Bunay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,T., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dachtorn,S.R., David,R.,  
 Delaney,K.R., Davis,C., Davy-Carroll,U., Deedrich,D.A.,  
 Delgado,K.J., Delgado,O., Denn,A.L., Ding,I., Dinh,H.H.,  
 Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Eathart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,L.E.,  
 Homel,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,B.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C.,  
 Kretovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgied,H.,  
 Lofado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maneswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
 Moser,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogun,M., Okunolu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Oulles,K., Ren,Y.,  
 Rivers,M., Rojka,A., Rojuben,I., Rolfe,M., Ruiz,S., Saverly,G.,  
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,  
 Sodavren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinsom,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
 Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G., and Gibbs,R.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 174028)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 3 (bases 1 to 174028)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission

JOURNAL COMMENT

Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 10, 2002 this sequence version replaced gi:17942431.  
 -----  
 Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 Project Information  
 Center project name: GDIS  
 Center clone name: CH230-901  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 121217 bases at least Q40  
 Consensus quality: 127332 bases at least Q30  
 Consensus quality: 131924 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length.  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 67 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 1 1004: contig of 1004 bp in length  
 1005 1104: gap of unknown length  
 1105 2301: contig of 1197 bp in length  
 2302 2401: gap of unknown length  
 2401 3456: contig of 1055 bp in length  
 3457 3556: gap of unknown length  
 3557 4594: contig of 1038 bp in length  
 4595 4694: gap of unknown length  
 4695 5999: contig of 1205 bp in length  
 5999 6000: gap of unknown length  
 6000 7391: contig of 1392 bp in length  
 7392 7491: gap of unknown length  
 7492 8563: contig of 1072 bp in length  
 8564 8663: gap of unknown length  
 8664 9676: contig of 1013 bp in length  
 9677 11543: gap of unknown length  
 11544 11643: contig of 1767 bp in length  
 11644 12808: gap of unknown length  
 12809 12908: contig of 1165 bp in length  
 12909 14320: gap of unknown length  
 14320 14421: contig of 1412 bp in length  
 14421 15913: gap of unknown length  
 15914 16013: contig of 1493 bp in length  
 16014 17145: gap of unknown length  
 17146 17245: contig of 1132 bp in length  
 17246 18550: gap of unknown length  
 18551 18650: contig of 1305 bp in length  
 18651 20070: gap of unknown length  
 20071 21206: contig of 1420 bp in length  
 21207 21306: gap of unknown length  
 21307 22658: contig of 1036 bp in length  
 22659 22758: gap of unknown length  
 22759 24735: contig of 1352 bp in length  
 24736 24835: gap of unknown length  
 24836 26165: gap of unknown length  
 26166 26265: contig of 1330 bp in length  
 26266 28440: gap of unknown length  
 28441 28540: contig of 2175 bp in length  
 28541 30485: gap of unknown length  
 30486 31765: contig of 1945 bp in length  
 31766 31765: gap of unknown length  
 31766 31765: contig of 1180 bp in length



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 18:55:48 ; Search time 351 Seconds  
(without alignments)  
9014.409 Million cell updates/sec

Title: US-09-820-003b-1

Perfect score: 1405  
Sequence: 1 aagcagatagctgagtcgagc.....aaaaaaaaaaaaaaaaaaaaa 1405

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N\_Genseq 101002.\*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1200.2	85.4	2528	21	AAC60009 Human secreted pro
2	1029	73.2	1193	23	AAS83860 DNA encoding novel
3	506	36.0	506	24	ABR44823 cDNA encoding novel
4	476.2	33.9	723	24	ABR83978 Human cDNA diffe
5	467	33.2	487	22	ABR43096 Human breast cell
6	467	33.2	487	22	ABR53512 Human foetal liver
7	467	33.2	487	22	ABR53512 Probe #1742 for ge
8	467	33.2	487	22	ABR53512 Human brain expres
9	467	33.2	487	22	ABR53512 Human bone marrow

10	467	33.2	487	22	AA118918	Probe #1751 for ge
11	467	33.2	487	22	AA133136	Probe #1822 used t
12	467	33.2	487	22	AA101752	Probe #1743 used t
13	467	33.2	487	21	ABSO1772	Human genome-deriv
14	333.4	23.7	335	21	AAC06621	Human secreted pro
15	264	18.8	406	23	ABV14061	Human prostate exp
16	264	18.8	448	23	ABV35165	Human prostate exp
17	264	18.6	448	23	ABV44007	Human prostate exp
18	261	18.6	200	24	ABL75210	Corn tassell-deriv
19	232.8	16.6	241	22	AAFL7849	Human breast cancer
20	232.8	16.6	241	22	AAFL7849	Human breast cancer
21	229	15.8	250	22	AAV04892	Human prostate exp
22	222.4	15.8	229	22	AAV88033	EST clone FC327.
23	219.8	15.6	229	22	AAH55559	Human breast tumor
24	195	13.9	925	20	AAK27232	Human Rab protein,
25	195	13.9	939	21	AAFL2161	Human breast and o
26	195	13.9	1898	22	AAK44924	Human polynucleoti
27	195	13.9	1944	22	AAK44924	DNA encoding novel
28	190.8	13.6	737	23	AAK3862	Human secreted exp
29	188.8	13.4	214	21	AAK42979	Human secreted exp
30	187	13.3	1965	22	AAK34668	Human secreted exp
31	179	12.7	179	22	ABR48222	Human secreted exp
32	179	12.7	179	22	ABR48222	Human secreted exp
33	179	12.7	179	22	ABR48222	Human secreted exp
34	179	12.7	179	22	ABR48222	Human secreted exp
35	179	12.7	179	22	ABR48222	Human secreted exp
36	179	12.7	179	22	ABR48222	Human secreted exp
37	179	12.7	179	22	ABR48222	Human secreted exp
38	179	12.7	179	22	ABR48222	Human secreted exp
39	179	12.7	179	22	ABR48222	Human secreted exp
40	175.2	12.5	185	16	ABR14259	Human genome-deriv
41	171	11.8	151	20	AAK19274	Human gene signatu
42	168	11.6	166	20	AAK23507	Human neutrophil c
43	163	11.6	1791	23	ABL05963	Human neutrophil c
44	155.4	11.1	607	22	AAH87926	Drosophila melanog
45	150.8	10.7	932	21	AAC35200	Peppermint plant o

#### ALIGNMENTS

RESULT 1	
AAC60009	Human secreted protein gene 44 SPQ ID NO:54.
ID	AAC60009 standard; CDNA; 2528 BP.
XX	
AC	AAC60009;
XX	
DT	26-JAN-2001 (first entry)
XX	
DE	Human secreted protein gene 44 SPQ ID NO:54.
XX	
KW	Human; secreted protein; diagnosis; cytosolic; immunosuppressive;
KW	neurotrophic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW	antidiabetic; antiinflammatory; antitumor; anticonvulsant;
KW	antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW	cancer; immune disorder; cardiovascular disorder; wound healing;
XX	neurological disease; infectious disease; chromosome identification; ss.
OS	Homo sapiens.
XX	
PN	WO200058356-A1.
XX	
PD	05-OCT-2000.
XX	
PF	22-MAR-2000; 2000WO-US07535.
XX	
PR	26-MAR-1999; 98US-0126511.
FR	17-DEC-1999; 98US-0172413.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM, Komatsoulis G;
XX	



KW	Human: chromosome mapping; gene mapping; gene therapy; forensic;
KV	food supplement; medical imaging; diagnostic; genetic disorders; ss.
XX	
OS	Homo sapiens.
PX	WO2001/75067-A2.
PD	11-OCT-2001.
PF	30-MAR-2001; 2001MO-US08631.
PR	31-MAR-2000; 2000US-0540217.
PP	23-APR-2000; 2000US-0649167.
PA	(HYSE)- HYSED INC.
PL	Dmanac RT, Liu C, Tang YT;
DR	WPI; 2001-639362/73.
XX	P-PsDB; ABG19673.
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
PS	
XX	Claim 1; SEQ ID NO 19664; 103bp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of states expressing (II). (I) and (II) are useful for treating
CC	diseases involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 1193 BP; 372 A; 279 C; 213 G; 329 T; 0 other;
Query Match	73.2%; Score 1029; DB 23; Length 1193;
Best Local Similarity	89.9%; Pred. No. 8.7e-132;
Matches 1182; Conservative	0; Mismatches 0; Indels 133; Gaps 2
OY	1 AAGCGATACCTACG- GCAGGGCAGTCGATGTGTTTCAAGGACGAGTAGGGGAAGA 59
Db	1183 AAGGATACCTACGAGGCGCGCGCTCGATGTGTTTCAAGGACGAGTAGGGGAAGA 112
OY	60 CGTTTCCTGCCCGAAGACACCATTGCATCTCTTTTGATTAACCGTGGCGGAG 119
Db	1123 CGTTTCCTGCCCGAAGACACCATTGCATCTCTTTTGATTAACCGTGGCGGAG 106
OY	120 AGTCAGGGGGCGGCTCGGCAACAAGGGCGCGGTGGGCGCGGCGGACGTGACTGAC 179
Db	1063 AGTCAGGGGGCGGCTCGGCAACAAGGGCGCGGTGGGCGCGGCGGACGTGACTGAC 100
OY	180 ATGTCCAGCATGATCCCGAAATGTATTATTATTCAGATTACTGTGATGGGACTCA 239
Db	1003 ATGTCCAGCATGATCCCGAAATGTATTATTATTCAGATTACTGTGATGGGACTCA 944
OY	240 GGGGTTGGAAAGCTTCCTCTCTTCTTATAGTTTGCAGATGATACATACAGAAAAGCTAC 299

[illegible]





SO Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.2%; Score 467; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 2,4e-55;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 502 AGAATTTGGTGATTCCTTGGAATTCGGTTTTGGAAACAGTCTAAGATGCACGA 561
DB 467 AGGAATTTGGTGATTCCTTGGAATTCGGTTTTGGAAACAGTCTAAGATGCACGA 408
OY 562 ATGTAGAAGACGCTTTCATGAGAGTGGAGTGGATTTAAAGAGATGGTCCCGAG 621
DB 407 ATGTAGAAGACGCTTTCATGAGAGTGGAGTGGATTTAAAGAGATGGTCCCGAG 348
OY 622 CAACAGCTGTGTGTGTGAGAGTCCATGTTAAATTCAGAGCACTCCAGTCAAGCAGT 681
DB 347 CAACAGCTGTGTGTGTGAGAGTCCATGTTAAATTCAGAGCACTCCAGTCAAGCAGT 288
OY 682 CAGGTGAGGTGTGTGTGCTAATAATTTGCTCCATTCCTTTTTCACAGCAATGATTTGCA 741
DB 287 CAGGTGAGGTGTGTGTGCTAATAATTTGCTCCATTCCTTTTTCACAGCAATGATTTGCA 228
OY 742 TCTGAACCCAGTGAATAAACAATAATGCTGAATTTGCTGATGTAGTGCACATACAA 801
DB 227 TCTGAACCCAGTGAATAAACAATAATGCTGAATTTGCTGATGTAGTGCACATACAA 168
OY 802 CAGATTTCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 861
DB 167 CAGATTTCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 108
OY 862 TTATTCCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 921
DB 107 TTATTCCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 48
OY 922 CTGTTTATTAATAATGCTGTATATCCTTGTGCTTCTGATATCC 968
DB 47 CTGTTTATTAATAATGCTGTATATCCTTGTGCTTCTGATATCC 1

```

#### RESULT 6

ABAS3512/c

ID ABA3512 standard; DNA; 487 BP.

XX ABA3512;

XX 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #1817.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 PS Claim 1; SEQ ID NO 1817; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC foetal liver. The present sequence is a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences.

SO Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.2%; Score 467; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 2,4e-55;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 502 AGAATTTGGTGATTCCTTGGAATTCGGTTTTGGAAACAGTCTAAGATGCACGA 561
DB 467 AGGAATTTGGTGATTCCTTGGAATTCGGTTTTGGAAACAGTCTAAGATGCACGA 408
OY 562 ATGTAGAAGACGCTTTCATGAGAGTGGAGTGGATTTAAAGAGATGGTCCCGAG 621
DB 407 ATGTAGAAGACGCTTTCATGAGAGTGGAGTGGATTTAAAGAGATGGTCCCGAG 348
OY 622 CAACAGCTGTGTGTGTGAGAGTCCATGTTAAATTCAGAGCACTCCAGTCAAGCAGT 681
DB 347 CAACAGCTGTGTGTGTGAGAGTCCATGTTAAATTCAGAGCACTCCAGTCAAGCAGT 288
OY 682 CAGGTGAGGTGTGTGTGCTAATAATTTGCTCCATTCCTTTTTCACAGCAATGATTTGCA 741
DB 287 CAGGTGAGGTGTGTGTGCTAATAATTTGCTCCATTCCTTTTTCACAGCAATGATTTGCA 228
OY 742 TCTGAACCCAGTGAATAAACAATAATGCTGAATTTGCTGATGTAGTGCACATACAA 801
DB 227 TCTGAACCCAGTGAATAAACAATAATGCTGAATTTGCTGATGTAGTGCACATACAA 168
OY 802 CAGATTTCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 861
DB 167 CAGATTTCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 108
OY 862 TTATTCCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 921
DB 107 TTATTCCTTACCGTCTCCACAAGGTCCAGAGATTGTAAATGGTCAATGACTTTT 48
OY 922 CTGTTTATTAATAATGCTGTATATCCTTGTGCTTCTGATATCC 968
DB 47 CTGTTTATTAATAATGCTGTATATCCTTGTGCTTCTGATATCC 1

```

#### RESULT 7

ABA23276/c

ID ABA23276 standard; DNA; 487 BP.

XX ABA23276;

XX 23-JAN-2002 (first entry)

DE Probe #1742 for gene expression analysis in human heart cell sample.

KW Human; gene expression; heart; microarray; vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease; ss.

XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.



XX 30-JAN-2001; 2001WO-US00666.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48889/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
XX Claim 1; SEQ ID No 1742; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart. The  
XX present sequence is one such probe. The probes may be used for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from the human heart via microarrays. By measuring gene expression, the  
XX probes are useful for predicting, diagnosing, grading, staging, and  
XX monitoring and prognosing diseases of the human heart and vascular system  
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
XX congenital heart disease.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;  
XX  
XX Query Match 33.2%; Score 467; DB 22; Length 487;  
XX Best Local Similarity 100.0%; Pred. No. 2.4e-55;  
XX Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 502 AGGAATTTGCTGATTCCTTGAATTCGTTTGGAAACCACTGCTAGATGCACAGA 561  
XX |||||||  
XX 467 AGGAATTTGCTGATTCCTTGAATTCGTTTGGAAACCACTGCTAGATGCACAGA 408  
XX |||||||  
XX 562 ATGTAGAACAAGTCTTTCATGACGATGGCAGCTGAGATTTAAAGCAATGGGTCGCGAG 621  
XX |||||||  
XX 407 ATGTAGAACAAGTCTTTCATGACGATGGCAGCTGAGATTTAAAGCAATGGGTCGCGAG 348  
XX |||||||  
XX 622 CAACAGCTGGTGTGCTGAGAAAGTCATGTTAAATTCAGACACTCCAGTCAAGCAGT 681  
XX |||||||  
XX 347 CAACAGCTGGTGTGCTGAGAAAGTCATGTTAAATTCAGACACTCCAGTCAAGCAGT 288  
XX |||||||  
XX 682 CAGGTGAGGTGCTGCTAAATTTGCTCCGCTCTCTTCTGCACAGCAATGATTTGCAA 741  
XX |||||||  
XX 287 CAGGTGAGGTGCTGCTAAATTTGCTCCGCTCTCTTCTGCACAGCAATGATTTGCAA 228  
XX |||||||  
XX 742 TCTGAACCAAGTGAAGAAAAACAAATTTGGCTCAATGCTAGTACGCTGCTGCAA 801  
XX |||||||  
XX 227 TCTGAACCAAGTGAAGAAAAACAAATTTGGCTCAATGCTAGTACGCTGCTGCAA 168  
XX |||||||  
XX 802 CAGATTTTACCGCTCTCCACAAAGTCAAGATTTGAATGCTCAATGCTGATTTT 861  
XX |||||||  
XX 167 CAGATTTTACCGCTCTCCACAAAGTCAAGATTTGAATGCTCAATGCTGATTTT 108  
XX |||||||  
XX 862 TTATTCCTTGAAGTCAAGACGCTAACTTCATTTTCAAGAACTGTTTAAACCTTTG 921  
XX |||||||  
XX 107 TTATTCCTTGAAGTCAAGACGCTAACTTCATTTTCAAGAACTGTTTAAACCTTTG 48  
XX |||||||  
XX 922 CTGCTTTTAAATAATATGTTGTAATCTGTTGCTTCCGTGATACC 968  
XX |||||||  
XX 47 CTGCTTTTAAATAATATGTTGTAATCTGTTGCTTCCGTGATACC 1

RESULT 8  
AAK01783/C  
ID AAK01783 standard; DNA; 487 BP.  
XX  
XX AAK01783;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 1774.  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 1774.  
XX  
XX Human: brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX  
XX Example 4; SEQ ID NO: 1774; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is one of the probes of the  
XX invention.  
XX  
XX Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;  
XX  
XX Query Match 33.2%; Score 467; DB 22; Length 487;  
XX Best Local Similarity 100.0%; Pred. No. 2.4e-55;  
XX Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 502 AGGAATTTGCTGATTCCTTGAATTCGTTTGGAAACCACTGCTAGATGCACAGA 561  
XX |||||||  
XX 467 AGGAATTTGCTGATTCCTTGAATTCGTTTGGAAACCACTGCTAGATGCACAGA 408  
XX |||||||  
XX 562 ATGTAGAACAAGTCTTTCATGACGATGGCAGCTGAGATTTAAAGCAATGGGTCGCGAG 621  
XX |||||||  
XX 407 ATGTAGAACAAGTCTTTCATGACGATGGCAGCTGAGATTTAAAGCAATGGGTCGCGAG 348  
XX |||||||  
XX 622 CAACAGCTGGTGTGCTGAGAAAGTCATGTTAAATTCAGACACTCCAGTCAAGCAGT 681  
XX |||||||  
XX 347 CAACAGCTGGTGTGCTGAGAAAGTCATGTTAAATTCAGACACTCCAGTCAAGCAGT 288  
XX |||||||  
XX 682 CAGGTGAGGTGCTGCTAAATTTGCTCCGCTCTCTTCTGCACAGCAATGATTTGCAA 741  
XX |||||||  
XX 287 CAGGTGAGGTGCTGCTAAATTTGCTCCGCTCTCTTCTGCACAGCAATGATTTGCAA 228  
XX |||||||  
XX 742 TCTGAACCAAGTGAAGAAAAACAAATTTGGCTCAATGCTAGTACGCTGCTGCAA 801

```

Db      227 TCTAACCAGTAAAAAACAATAATGCTGAATGTACTGTATGTAAGTCACACTACAA 168
Oy      802 CAGATTTCTACCGTCTCCACAAGAGTCAAGATTTGTAATGTCACATACCTTTT 861
Db      167 CAGATTTCTACCGTCTCCACAAGAGTCAAGATTTGTAATGTCACATACCTTTT 108
Oy      862 TTATTCCTTGACTCAAGACAGTAACTTCAATTTTCAGAACTGTTTAAACCTTTG 921
Db      107 TTATTCCTTGACTCAAGACAGTAACTTCAATTTTCAGAACTGTTTAAACCTTTG 48
Oy      922 CTGCTTTATAAATAATGCTGAATGCTGTAATGCTGCTTCTGATACC 968
Db      47 CTGCTTTATAAATAATGCTGAATGCTGTAATGCTGCTTCTGATACC 1

```

RESULT 9  
AAK27239/C  
ID AAK27239 standard; DNA; 487 BP.

AAK27239;

06-NOV-2001 (first entry)

Human bone marrow expressed single exon probe SEQ ID NO: 1796.

Human; bone marrow expressed exon; gene expression analysis; probe;  
microarray; cancer; leukemia; lymphoma; myeloma; ss.

Homo sapiens.

WO200157276-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00668.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human bone marrow -

Example 4; SEQ ID NO: 1796; 658bp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid

probes which are derived from genomic sequences expressed in the human

bone marrow. They can be used to measure gene expression in bone marrow

samples, which may enable the improved diagnosis and treatment of cancers

such as lymphoma, leukemia and myeloma. The present sequence is one of

the probes of the invention.

Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.2%; Score 467; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 2.4e-55;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 502 AGGAATTTGCTGATTCCTTGGAATTCGTTTGGAAACCAAGTCTAAGATGCAACA 561

Db 467 AGGAATTTGCTGATTCCTTGGAATTCGTTTGGAAACCAAGTCTAAGATGCAACA 408

```

Oy      562 ATGTAGAACAAGCTCTTTCATGACGATGGCAGCTGAGATTTAAAAAGCAATGGTCCGAG 621
Db      407 ATGTAGAACAAGCTCTTTCATGACGATGGCAGCTGAGATTTAAAAAGCAATGGTCCGAG 348
Oy      622 CAACAGCTGGTGGTCTGAGAAAGTCCAAATGTTAAATTCAGAGACTCCAGTCAAGAGT 681
Db      347 CAACAGCTGGTGGTCTGAGAAAGTCCAAATGTTAAATTCAGAGACTCCAGTCAAGAGT 288
Oy      682 CAGGTGAGAGTGGTGTCTAAATTTGCTCAGTCCATCTTTTTCAGAGCAATGATTTGCA 741
Db      287 CAGGTGAGAGTGGTGTCTAAATTTGCTCAGTCCATCTTTTTCAGAGCAATGATTTGCA 228
Oy      742 TCTGAACCAAGTGAATAAATAATGCTGAAATGCTGATGTAAGTGTAGCTGCACAA 801
Db      227 TCTGAACCAAGTGAATAAATAATGCTGAAATGCTGATGTAAGTGTAGCTGCACAA 168
Oy      802 CAGATTCCTACCGTCTCCACAAGAGTCAAGATTTGTAATGTCACATACCTTTT 861
Db      167 CAGATTCCTACCGTCTCCACAAGAGTCAAGATTTGTAATGTCACATACCTTTT 108
Oy      862 TTATTCCTTGACTCAAGACAGTAACTTCAATTTTCAGAACTGTTTAAACCTTTG 921
Db      107 TTATTCCTTGACTCAAGACAGTAACTTCAATTTTCAGAACTGTTTAAACCTTTG 48
Oy      922 CTGCTTTATAAATAATGCTGAATGCTGTAATGCTGCTTCTGATACC 968
Db      47 CTGCTTTATAAATAATGCTGAATGCTGTAATGCTGCTTCTGATACC 1

```

RESULT 10  
AA11818/C  
ID AA11818 standard; DNA; 487 BP.

AA11818;

12-OCT-2001 (first entry)

Probe #1751 for gene expression analysis in human cervical cell sample.

Probe; human; microarray; gene expression; cervical epithelial cell;

cervical cancer; ss.

Homo sapiens.

WO200157278-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00670.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID NO 1751; 487bp; English.

The present invention relates to human single exon nucleic acid probes

(SNPs). The present sequence is one such probe. The SNPs are derived

from human HeLa cells. The SNPs can be used to produce a single exon

microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: the sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).

XX Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.28; Score 467; DB 22; Length 487;  
 Best Local Similarity 100.0%; Pred. No. 2,4e-55;  
 Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 502 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCGATGCTAAGAAATGCAACGA 561
DB 467 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCGATGCTAAGAAATGCAACGA 408
OY 562 ATGTAGAACAGTCTTTCATGACGATGCGACGAGATTAAACCAATGGTCCCGAG 621
DB 407 ATGTAGAACAGTCTTTCATGACGATGCGACGAGATTAAACCAATGGTCCCGAG 348
OY 622 CACACAGCTGGTGGTGGTGGAGAGTCACATGTTTAAATTCAGAGCAGCTCCAGCAGT 681
DB 347 CACACAGCTGGTGGTGGTGGAGAGTCACATGTTTAAATTCAGAGCAGCTCCAGCAGT 288
OY 682 CAGGTGGAGGTGCTCTTAAATTTGCTCCATCCCTTTTCACAGCAATGAAATGGCAA 741
DB 287 CAGGTGGAGGTGCTCTTAAATTTGCTCCATCCCTTTTCACAGCAATGAAATGGCAA 228
OY 742 TCTGACCCCAAGTGAAGAAAAAAGAAATTCCTGATTTGATGATGATGATGATGATGAT 801
DB 227 TCTGACCCCAAGTGAAGAAAAAAGAAATTCCTGATTTGATGATGATGATGATGATGAT 168
OY 802 CAGATCTTACGCTCCCAAGAGTCAAGAGTGTAAATTCGCAATGCTGCTTTT 861
DB 167 CAGATCTTACGCTCCCAAGAGTCAAGAGTGTAAATTCGCAATGCTGCTTTT 108
OY 862 TTAATCCCTGATCAAGACGTAATCTTCAATTCACAAAGTGTAAACCTTTGTGTG 921
DB 107 TTAATCCCTGATCAAGACGTAATCTTCAATTCACAAAGTGTAAACCTTTGTGTG 48
OY 922 CTGGTTTAAATAATATGATGCTTAAATTCCTGCTTTCCTGATAC 968
DB 47 CTGGTTTAAATAATATGATGCTTAAATTCCTGCTTTCCTGATAC 1

```

RESULT 11  
 AAI3136/c  
 ID AAI3136 standard; DNA; 487 BP.

XX AAI3136;

DB 17-OCT-2001 (first entry)

DE Probe #1822 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

OS W0200157272-AZ.

XX W0200157272-AZ.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236559.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 1822; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SNP).

XX CC The present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.28; Score 467; DB 22; Length 487;  
 Best Local Similarity 100.0%; Pred. No. 2,4e-55;  
 Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 502 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCGATGCTAAGAAATGCAACGA 561
DB 467 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCGATGCTAAGAAATGCAACGA 408
OY 562 ATGTAGAACAGTCTTTCATGACGATGCGACGAGATTAAACCAATGGTCCCGAG 621
DB 407 ATGTAGAACAGTCTTTCATGACGATGCGACGAGATTAAACCAATGGTCCCGAG 348
OY 622 CACACAGCTGGTGGTGGTGGAGAGTCACATGTTTAAATTCAGAGCAGCTCCAGCAGT 681
DB 347 CACACAGCTGGTGGTGGTGGAGAGTCACATGTTTAAATTCAGAGCAGCTCCAGCAGT 288
OY 682 CAGGTGGAGGTGCTCTTAAATTTGCTCCATCCCTTTTCACAGCAATGAAATGGCAA 741
DB 287 CAGGTGGAGGTGCTCTTAAATTTGCTCCATCCCTTTTCACAGCAATGAAATGGCAA 228
OY 742 TCTGACCCCAAGTGAAGAAAAAAGAAATTCCTGATTTGATGATGATGATGATGATGAT 801
DB 227 TCTGACCCCAAGTGAAGAAAAAAGAAATTCCTGATTTGATGATGATGATGATGATGAT 168
OY 802 CAGATCTTACGCTCCCAAGAGTCAAGAGTGTAAATTCGCAATGCTGCTTTT 861
DB 167 CAGATCTTACGCTCCCAAGAGTCAAGAGTGTAAATTCGCAATGCTGCTTTT 108
OY 862 TTAATCCCTGATCAAGACGTAATCTTCAATTCACAAAGTGTAAACCTTTGTGTG 921
DB 107 TTAATCCCTGATCAAGACGTAATCTTCAATTCACAAAGTGTAAACCTTTGTGTG 48
OY 922 CTGGTTTAAATAATATGATGCTTAAATTCCTGCTTTCCTGATAC 968
DB 47 CTGGTTTAAATAATATGATGCTTAAATTCCTGCTTTCCTGATAC 1

```

RESULT 12  
 AAI01752/c  
 ID AAI01752 standard; DNA; 487 BP.

XX AAI01752;

DB 09-OCT-2001 (first entry)

DE Probe #1743 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

OS AAI01752;

PN WO200157270-A2.  
 XX 09-AUG-2001.  
 XX 29-JAN-2001; 2001WO-US00661.  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 XX in a human breast -  
 XX Claim 25; SEQ ID No 1743; 322pp; English.  
 XX The present invention relates to novel single exon nucleic acid probes.  
 XX The present sequence is one such probe. The probes are useful for  
 XX measuring human gene expression in a human breast sample, where the probe  
 XX hybridises at high stringency to a nucleic acid expressed in the human  
 XX breast. The probes are useful for predicting, diagnosing, grading,  
 XX staging, monitoring and prognosing diseases of the human breast,  
 XX particularly those diseases with polygenic aetiology. The diseases  
 XX include: breast cancer, disorders of development, inflammatory diseases  
 XX of the breast, fibrocystic changes, proliferative breast disease and  
 XX non-carcinoma tumours.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pcr\_sequences.  
 XX Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;  
 SQ  
 Query Match 33.2%; Score 467; DB 22; Length 487;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-55;  
 Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 502 AGGAATTGCTGATTCCTTGGAAATTCCTTTTGGAAACAGTGTAAAGTCAACGA 561  
 DB 467 AGGAATTGCTGATTCCTTGGAAATTCCTTTTGGAAACAGTGTAAAGTCAACGA 408  
 QY 562 ATGTGGAACAGCTTTCATGACGATGCGACGTGAGATTAAAAAGCAATGGGTCCGGAG 621  
 DB 407 ATGTGGAACAGCTTTCATGACGATGCGACGTGAGATTAAAAAGCAATGGGTCCGGAG 348  
 QY 622 CAACAGCTGCTGCTGCTGAGAGTCAATGTTAAATTCAGAGCACTCCAGTCAAGCACT 681  
 DB 347 CAACAGCTGCTGCTGCTGAGAGTCAATGTTAAATTCAGAGCACTCCAGTCAAGCACT 288  
 QY 682 CAGGTGAGAGTGTCTGCTAAATTTGCTCCATCTCTTCTCAGACGATGATTTGCA 741  
 DB 287 CAGGTGAGAGTGTCTGCTAAATTTGCTCCATCTCTTCTCAGACGATGATTTGCA 228  
 QY 742 TGTGAACCAAGTGAAGAAACAAATTTGCTGATTTGATGATGATGATGATGATGAT 801  
 DB 227 TGTGAACCAAGTGAAGAAACAAATTTGCTGATTTGATGATGATGATGATGATGAT 168  
 QY 802 CAGATTCTTACGCTCCCAAGAGTGAATTTGATGATGATGATGATGATGATGATGAT 861  
 DB 167 CAGATTCTTACGCTCCCAAGAGTGAATTTGATGATGATGATGATGATGATGATGAT 108  
 QY 862 TTATTCCTTGAAGTCAAGACGATGATTTTCAAGACGATTTTAAACCTTTGTGTG 921  
 DB 107 TTATTCCTTGAAGTCAAGACGATGATTTTCAAGACGATTTTAAACCTTTGTGTG 48

QY 922 CTGTTTAAATAATATGTTGATTCCTTGTGCTTCCGTGATACC 968  
 DB 47 CTGTTTAAATAATATGTTGATTCCTTGTGCTTCCGTGATACC 1  
 RESULT 13  
 ABS01772/c  
 ID ABS01772 standard; DNA; 487 BP.  
 XX ABS01772;  
 AC 19-AUG-2002 (first entry)  
 DE Human genome-derived single exon probe from lung SEQ ID No 1763.  
 XX Human: ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX Homo sapiens.  
 OS WO200186003-A2.  
 PN 15-NOV-2001.  
 PD 30-JAN-2001; 2001WO-US00665.  
 PF 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 XX measure gene expression in human lung samples -  
 XX Claim 1; SEQ ID No 1763; 634pp; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 XX nucleic acid probes for measuring gene expression in a sample derived  
 XX from human lung comprising single exon nucleic acid probes having one of  
 XX 12614 nucleic acid sequences mentioned in the specification, or their  
 XX complements or the 12387 open reading frames derived from the 12614  
 XX probes. Also included are a microarray comprising the novel set of  
 XX probes; the novel set of probes which hybridise at high stringency to a  
 XX nucleic acid expressed in the human lung; measuring gene expression in a  
 XX sample derived from human lung, comprising (a) contacting the array with  
 XX a collection of detectably labeled nucleic acids derived from human lung  
 XX mRNA, and (b) measuring the labeled nucleic acids bound to each probe of  
 XX the array; identifying exons in a eukaryotic genome, comprising  
 XX (a) algorithmically predicting at least one exon from genomic sequences  
 XX of the eukaryote; and (b) detecting specific hybridisation of detectably  
 XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 XX having a fragment identical to the predicted exon, the probe is included  
 XX in the above mentioned microarray; assigning exons to a single gene,  
 XX comprising (a) identifying exons from genomic sequence by the method  
 XX above and (b) measuring the expression of each of the exons in several  
 XX tissues and/or cell types using hybridisation to a single exon  
 XX microarrays having a probe with the exon, where a common pattern of

CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences; mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagenen syndrome, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a single exon  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIP0 at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.2%; Score 467; DB 24; Length 487;

Best Local Similarity 100.0%; Pred. No. 2.4e-55; Mismatches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 502 AGGAATTTGGTATCCCTTGAATTCCTTTTGAACAGAGGTGTAAGATGCAACA 561  
|||||  
DB 467 AGGAATTTGGTATCCCTTGAATTCCTTTTGAACAGAGGTGTAAGATGCAACA 408  
|||||  
DB 562 ATGAGAACAGCTTTTCATGAGCATGGAGCTGATTAAGAACAGATGGTCCGGAG 621  
|||||  
DB 407 ATGAGAACAGCTTTTCATGAGCATGGAGCTGATTAAGAACAGATGGTCCGGAG 348  
|||||  
DB 622 CAACAGCTGGTGGTGTGAGAGAGTCAATGTAATTAATGACAGCTCCAGCAAGCAT 661  
|||||  
DB 347 CAACAGCTGGTGGTGTGAGAGAGTCAATGTAATTAATGACAGCTCCAGCAAGCAT 288  
|||||  
DB 682 CAGGTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 741  
|||||  
DB 287 CAGGTGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 228  
|||||  
DB 742 TCTGACCCCAAGTGAACCAAAATGGCTGAATTTGATGTAAGTCTGACTACAA 801  
|||||  
DB 227 TCTGACCCCAAGTGAACCAAAATGGCTGAATTTGATGTAAGTCTGACTACAA 168  
|||||  
DB 802 CAGATTTTACGCTGCCCAAGGTCAGATTTGTAATGGTCAATGACTGATTTTTT 861  
|||||  
DB 167 CAGATTTTACGCTGCCCAAGGTCAGATTTGTAATGGTCAATGACTGATTTTTT 108  
|||||  
DB 862 TTAATCCCTGACCTCAAGACAGTCAATTTTCAAGATGTTTAAACCTTTGCTG 921  
|||||  
DB 107 TTAATCCCTGACCTCAAGACAGTCAATTTTCAAGATGTTTAAACCTTTGCTG 48  
|||||  
DB 922 CTGGTTATTAATAATATGTGTATTCCTTTGCTTTCCGATAC 968  
|||||  
DB 47 CTGGTTATTAATAATATGTGTATTCCTTTGCTTTCCGATAC 1

## RESULT 14

AA06621 standard; cDNA; 335 BP.

AA06621;

06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 10696.  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
XX Homo sapiens.

XX EP1033401-A2.  
PN 06-SEP-2000.  
XX 21-FEB-2000; 2000EP-0200610.  
XX 26-FEB-1999; 99US-0122487.  
XX (GSEST) GENSET.  
XX Dunas Milne Edwards J, Duclert A, Giordano J;  
XX WPI: 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 10696; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
CC cDNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

SO Sequence 335 BP; 96 A; 70 C; 56 G; 109 T; 4 other;

Query Match 23.7%; Score 333.4; DB 21; Length 335;

Best Local Similarity 98.8%; Pred. No. 3.6e-37; Mismatches 331; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

DB 646 CCATGTTTAAATTCAGACACTCCAGTCAAGCAGTCAAGTGGAGTTGCTTAATTT 705  
|||||  
DB 1 CCATGTTTAAATTCAGACACTCCAGTCAAGCAGTCAAGTGGAGTTGCTTAATTT 60  
|||||  
DB 706 TGCTCCATCTCTTTTCACAGCAATGTAATTTGCAATGTAATTTGCAATTTG 765  
|||||  
DB 61 TGCTCCATCTCTTTTCACAGCAATGTAATTTGCAATGTAATTTGCAATTTG 120  
|||||  
DB 766 ATTGCGTAATTTGAGTGTATGTAGTCACTACACAGATTTTACCGTCCCAAG 825  
|||||  
DB 121 ATTGCGTAATTTGAGTGTATGTAGTCACTACACAGATTTTACCGTCCCAAG 180  
|||||  
DB 826 GTGAGAGTGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 885  
|||||  
DB 181 GTGAGAGTGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
|||||  
DB 886 AACTTCATTTTCAAGACAGTGTAAACCTTTGCTTTGCTTTGCTTTGCTTTGCTTT 945  
|||||  
DB 241 AACTTCATTTTCAAGACAGTGTAAACCTTTGCTTTGCTTTGCTTTGCTTTGCTTT 300  
|||||  
DB 946 ATCCGTGTGCTTTTCTGATACAGAGCTGTTCC 980  
|||||  
DB 301 ATCCGTGTGCTTTTCTGATACAGAGAGATGTTCC 335

## RESULT 15

ABV14061/c standard; cDNA; 406 BP.

ABV14061;

XX

DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker CDNA 14052.  
 XX  
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KM pharmacogenomic marker; gene; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-0505171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 2343-2344; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 CC  
 XX  
 SQ Sequence 406 BP; 104 A; 99 C; 92 G; 111 T; 0 other;  
 XX  
 Query Match 18.8%; Score 264; DB 23; Length 406;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-28;  
 Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 518 CCTGGAATTCGTTTGGAAACGAGTCTAGATGCAACGAATGTAGAACAGTCTTT 577  
 DB 406 CCTGGAATTCGTTTGGAAACGAGTCTAGATGCAACGAATGTAGAACAGTCTTT 347  
 QY 578 CATGACATGCGAGCTGAGATTAAAGGAAATGGTCCCGAGACACAGCTGGTGTGC 637  
 DB 346 CATGACATGCGAGCTGAGATTAAAGGAAATGGTCCCGAGACACAGCTGGTGTGC 287  
 QY 638 TGAGAGTCCATGTAAATCAAGACATCCAGTCAAGTCAAGTGAAGTGTGCTG 697  
 DB 286 TGAGAGTCCATGTAAATCAAGACATCCAGTCAAGTCAAGTGAAGTGTGCTG 227  
 QY 698 CTAATATTTGCTCCATCTTTCTACAGCAATGAATTTGCAATCTGAACCAAGTGAA 757  
 DB 226 CTAATATTTGCTCCATCTTTCTACAGCAATGAATTTGCAATCTGAACCAAGTGAA 167  
 QY 758 AAAACAAATTCCTGAATGTAC 781

Db |||||  
 166 AAAACAAATTCCTGAATGTAC 143  
 Search completed: June 22, 2003, 19:53:56  
 Job time : 356 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 19:33:58 ; Search time 1902 Seconds  
(without alignments)  
11963.543 Million cell updates/sec

Title: US-09-820-003b-1  
Perfect score: 1405  
Sequence: 1 aagcagatgctgagtcgcgc.....aaaaaaaaaaaaaaaaaa 1405

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estlun.\*  
5: em\_estlov.\*  
6: em\_estlpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estcom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_hiv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_trod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	743	52.9	948	9	AL539022
2	732.4	52.1	777	14	BQ014597
3	727.6	51.8	811	13	AA746643
4	723.8	51.5	817	13	BI869965
5	708.8	50.4	910	14	BQ276678
6	708.4	50.4	761	14	BQ772048

7	707.8	50.4	945	9	AL530265
8	704.4	50.1	1105	13	BM452262
9	703.6	50.1	726	14	BQ448090
10	702	50.1	766	14	BI913092
11	692.8	49.3	760	14	BM971301
12	691.8	49.2	976	14	BM912860
13	690.6	49.2	837	13	BI086445
14	689	49.0	689	14	BM706159
15	688.2	49.0	737	13	BI092003
16	686	48.8	686	14	BM826571
17	685.6	48.8	717	9	AI952757
18	683.2	48.6	706	14	BM975245
19	682.6	48.2	796	14	BI659520
20	677.6	48.2	877	12	BG617277
21	676.4	48.1	678	13	BG939358
22	671.2	47.8	684	13	BI495589
23	670.8	47.7	1130	13	BM542833
24	669.6	47.7	813	14	BQ179586
25	668.2	47.6	812	14	BQ178972
26	661.2	47.1	686	14	BQ599300
27	659.6	46.9	721	9	AI884351
28	658	46.8	793	12	BG178268
29	657.6	46.8	690	10	AM150848
30	655.2	46.6	692	14	BQ010358
31	650.4	46.3	692	9	AI828874
32	650	46.3	908	9	AA659410
33	648	46.1	723	9	AI800097
34	645.6	46.0	682	9	AI858694
35	640.4	45.6	900	12	BG032817
36	638.8	45.5	682	9	AI860577
37	638	45.4	638	14	BM783973
38	638	45.4	646	14	BQ599347
39	638	45.4	687	10	BE646347
40	637.4	45.4	719	14	BQ210709
41	632.8	45.0	1055	13	BM472577
42	631.8	45.0	664	10	BM617921
43	631.2	44.9	862	14	BQ223273
44	630.6	44.9	645	10	AM166902
45	628.6	44.7	702	9	AI89183

#### ALIGNMENTS

RESULT 1  
AL539022  
LOCUS AL539022 948 bp mRNA linear EST 16-FEB-2001  
DEFINITION LTI\_FL013.FBrl Homo sapiens cDNA clone CS0DF030YG06 5  
ACCESSION AL539022  
VERSION AL539022.1 GI:12867866  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
source  
1. 948  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DF030YG06"  
/clone\_1bp="LTI\_FL013.FBrl"  
/dev\_stage="pooled tissue from post conception fetuses (20 week - stage and 26 week)"  
/lab\_host="DH10B"







Oy	1162	GGTGAGATTAAATGGAAAATACTACACTCGTGAATTAATGTTCTTCATACCTGCAT	1221
.Db	154	TGTGAGATTAAATGGAAAATACTACACTCACTCGAATTAATGTTCTTCATACCTGCAT	95
Oy	1222	ATAATTTGTGCTGCAGAAATATTGTAATTTGTGCACACTANGTAACAAACAACCTGAAG	1281
.Db	94	ATAATTTGTGCTGCAGAAATATTGTAATTTGTGCACACTANGTAACAAACAACCTGAAG	35
Oy	1282	ATAATGTTTAATAATATATGTACTATTGGAAGA	1315
.Db	34	ATAATGTTTAATAATATATGTACTATTGGAAGA	1

RESULT 4	817 bp	mRNA	linear	EST 11-OCT-2001
B1869965	603394116p1	NIH_MGC_90	Homo sapiens	CDNA IMAGE:5404275 5',
LOCUS	B1869965			
DEFINITION	mRNA sequence.			
ACCESSION	B1869965			
VERSION	B1869965.1			
KEYWORDS	GI:16043638			
SOURCE	EST.			
	human.			

REFERENCE	1 (bases 1 to 817)
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.jnl.gov>  
 Plate: IMAGE2030 Row: p Column: 04  
 High quality sequence stop: 669.  
 Location/Qualifiers

```

/ad_xref="taxon:9606"
/clone_image:5404275"
/clone_1b="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10 (phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."

```

Query Match	51.5%	Score 723.8	DB 13	Length 817
Best Local Similarity	95.3%	Pred. No. 2.7e-76		
Matches 779	Conservative 0	Mismatches 32	Indels 6	Gaps 3
Qy	475	AGAAAGTGTAGTACACACACAGGAAAGTAATTTGGATTCCCTTGGAAATTCGGTTT	534	
Db	1	AGAAAGTGTAGTACACACACAGGAAATTTGGATTCCCTTGGAAATTCGGTTT	60	
Qy	535	TGGAACCAAGTCTAGAGATGCAACGATGTAGAACACTCTTTCATGACATGGCAGCTG	594	
Db	61	TGGAACCAAGTCTAGAGATGCAACGATGTAGAACACTCTTTCATGACATGGCAGCTG	120	
Qy	595	AGATTAAAACCGAATGGGTCCCGAGACACAGCTGGTGGTGTAGAAAGTCCAAATGTTA	654	
Db	121	AGATTAAAACCGAATGGGTCCCGAGACACAGCTGGTGGTGTAGAAAGTCCAAATGTTA	180	

OY	655	AAATTCAGAGACATCCAGTCAAGACAGTCAGGTGGTGGTGGTAAATTTGGCTCAT	714
Db	181	AAATTCAGAGACATCCAGTCAAGACAGTCAGGTGGTGGTGGTAAATTTGGCTCAT	240
OY	715	CCCTTTCTCAGACGAATGAATTTGCCAATCTGAAACCCAAAGTGAATAAAATTTGCCCTGA	774
Db	241	CCCTTTCTCAGACGAATGAATTTGCCAATCTGAAACCCAAAGTGAATAAAATTTGCCCTGA	300
OY	775	ATTGTACTGTATGTAGCTGCACCTACACAGATTTCTTACCGTCTCCACAAAGGTACAGAT	834
Db	301	ATTGTACTGTATGTAGCTGCACCTACACAGATTTCTTACCGTCTCCACAAAGGTACAGAT	360
OY	835	TGTAAATGGTCAATACTGACTTTTATTTTATTTCCCTGACACACAGACGTAACTTAT	894
Db	361	TGTAAATGGTCAATACTGACTTTTATTTTATTTCCCTGACACACAGACGTAACTTAT	420
OY	895	TTTCAGAACTGTTTAAACCTTTGATGCTGGTTATATAAATAATGTGTATACCTGTT	954
Db	421	TTTCAGAACTGTTTAAACCTTTGATGCTGGTTATATAAATAATGTGTATACCTGTT	480
OY	955	GCTTTCGATACACAGACTGTTTCCCGGGTGGTTGAATATATTTGGTTTGAATGTTT	1014
Db	481	GCTTTCGATACACAGACTGTTTCCCGGGTGGTTGAATATATTTGGTTTGAATGTTT	540
OY	1015	ATATTTGGCATTTAGATGTGCAGGTTTAGTCTTGAAAGATGAAGTTCAGCATTGTTGTA	1074
Db	541	ATATTTGGCATTTAGATGTGCAGGTTTAGTCTTGAAAGATGAAGTTCAGCATTGTTGTA	600
OY	1075	TCAAAACAGCACAAACAGATGTGTGCACCTTTCAGATGCATTAAGTTTAACTAGATGTTTAT	1134
Db	601	TCAAAACAGCACAAACAGATGTGTGCACCTTTCAGATGCATTAAGTTTAACTAGATGTTTAT	660
OY	1135	GTAAGATCT---GATTTGCTAGTCTTCCCTGTGAGAGTTATTAATGGAAGATTAACCTA	1191
Db	661	GTAAGATCTGATTTGGCTAGTCTTCCCTGTGAGAGTCATTAATGGAAGATTAACCTA	720
OY	1192	TCTGATTAATAAGT-TTCTTCATACTGTCATATAATTGTGGCTGCAGAA--TATTGTAA	1248
Db	721	TCTGATTAATAAGTCTCTTCATTACTTCGCAATATCATCGGGGCTGCAGAACTCTTGCAA	780
OY	1249	TTTGTTCGACACTGTATGAACAAACAACTGAAGAT	1285
Db	781	TATGTTCGCCACTATGTACAAAACAACTGCAGAGTAT	817

FEATURES	RESULT 5
LOCUS	BQ276678
DEFINITION	BQ276678 910 bp mRNA linear EST 07-MAY-2002
ACCESSION	AGNCOURT_7025320 NIH_MGC_126 Homo sapiens cDNA clone IMAGE:5809629
VERSION	5, mRNA sequence.
KEYWORDS	BQ276678
SOURCE	BQ276678.1 GI:20486886
ORGANISM	EST.
REFERENCE	human.
TITLE	Homo sapiens
AUTHORS	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
COMMENT	1 (bases 1 to 910)
	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
	National Institutes of Health, Mammalian Gene Collection (MGC)
	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>
	Tissue Procurement: NCI
	cDNA Library Preparation: Michael Brownstein Laboratory
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
	DNA Sequencing by: Agencourt Bioscience Corporation
	Clone distribution: MGC clone distribution Information can be
	found through the I.M.A.G.E.. Consortium/LNL at:
	<a href="http://image.lnl.gov">http://image.lnl.gov</a>
	Plate: LCM205 row: 1 column: 22
	High quality sequence stop: 558.
	location/Qualifiers

```

source
1. .910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5809629"
/clone_1lb="NIH_MGC_126"
/tissue="type-mixed (pool of 40 RNAs)"
/lab_host="DH10B (TI-phage-resistant)"
/note="Vector: pDR-LIB; Site_1: SfiI (ggccattatgccc);
Site_2: SfiI (ggccgcctggccc); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon
- 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
used in cloning as follows:
5'-AGCAGTGTATCAACGACGAGTGGCCATTCAGCGCGG-3' and
5'-ATTCTAGAGCCGCGGCGGCGCATG-dt(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5-1
kb size fraction (other fractions present in NIH_MGC_127
and NIH_MGC_128). Library created in the laboratory of T.
Uedlin, M.D., Ph.D. (NIH). Note: this is a NIH_MGC
Library."
BASE COUNT      243 a      194 c      180 g      292 t      1 others
ORIGIN
Query Match      50.4%; Score 708.8; DB 14; Length 910;
Best Local Similarity 98.9%; Pred. No. 1.4e-74;
Matches 724; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

OY 584 GATGGCAGCGATTAATAAAGCAATGGTCCCGAGCAACGCTGCGCTGAGAA 643
DB 3 GATGGCAGCGATTAATAAAGCAATGGTCCCGAGCAACGCTGCGCTGAGAA 62
OY 644 GTCCAGTAAATTAATGACAGCATCTCCAGTCAAGAGTGGAGTCTGCTTAAA 703
DB 63 GTCCAGTAAATTAATGACAGCATCTCCAGTCAAGAGTGGAGTCTGCTTAAA 122
OY 704 TTTCGCTCCATCTTTCTCAGACAGTAATTTGCAATCTGAACCCAGTAAAAA 763
DB 123 TTTCGCTCCATCTTTCTCAGACAGTAATTTGCAATCTGAACCCAGTAAAAA 182
OY 764 AAATTCGCTAAATTTGATGATGTAGTGTGACACTCAACAGATTTCCGCTCCAAA 823
DB 183 AAATTCGCTAAATTTGATGATGTAGTGTGACACTCAACAGATTTCCGCTCCAAA 242
OY 824 AGGTGACAGATTTGTAATGTGTAATGACTTTTATTTTATTCCTGACTCAAGAC 883
DB 243 AGGTGACAGATTTGTAATGTGTAATGACTTTTATTTTATTCCTGACTCAAGAC 302
OY 884 CTAACTTCATTTTCAAGACTGTTTAAACCTTGTGCTGCTGTTTAAATATGCTG 943
DB 303 CTAACTTCATTTTCAAGACTGTTTAAACCTTGTGCTGCTGTTTAAATATGCTG 362
OY 944 TTAATCTTGTGCTTCCGTGATACAGACGTTTCCGCTGCTGTTAGATAATATTTTG 1003
DB 363 TTAATCTTGTGCTTCCGTGATACAGACGTTTCCGCTGCTGTTAGATAATATTTTG 422
OY 1004 TTTTGAATTTAATTTGAGATGTTAGATGAGTTAGTCTTGAAGATGAGTTCA 1063
DB 423 TTTTGAATTTAATTTGAGATGTTAGATGAGTTAGTCTTGAAGATGAGTTCA 482
OY 1064 GCATTTTGTATCAACAGACAGAGTGTGCTGCACTTCCATGATATAAAGTTATG 1123
DB 483 GCATTTTGTATCAACAGACAGAGTGTGCTGCACTTCCATGATATAAAGTTATG 542
OY 1124 AGATGTTAATTTGTAAGATCTGATTTGCTAGTCTTCCCTTGAAGTTAATAATGGAAGA 1183
DB 543 AGATGTTAATTTGTAAGATCTGATTTGCTAGTCTTCCCTTGAAGTTAATAATGGAAGA 602
OY 1184 TTACACTCTCTGATTAATAGTTTCTTCAATCTGCAATATAATTTGCTGCAATAT 1243

```

```

DB 603 TTACACTCTCTGATTAATAGTTTCTTCAATCTGCAATATAATTTGCTGCAATAT 662
OY 1244 TGTAAATTTGTCGACACTGATGTACAAAACACTGAAGATA-TGTTAATAAATTTGTA 1302
DB 663 TGTAAATTTGTCGACACTGATGTACAAAACACTGAAGATA-TGTTAATAAATTTGTA 722
OY 1303 CTTATTTGGAAGT 1314
DB 723 CTTATTTGGAAGT 734

RESULT 6
BO772048/C 761 bp mRNA linear EST 26-JUL-2002
LOCUS BO772048
DEFINITION UT-H-EZ1-bdk-1-10-0-UI.s1 NCI-CGAP-Ch2 Homo sapiens cDNA clone
BO772048 UT-H-EZ1-bdk-1-10-0-UI 3', mRNA sequence.
ACCESSION BO772048.1 GI:21980524
VERSION BO772048
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 761)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Steven Citelis/ Rush Presbyterian, Dept. of
Orthopaedics
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA-Tes.

FEATURES
Location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UT-H-EZ1-bdk-1-10-0-UI"
/clone_1lb="NCI-CGAP-Ch2"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pelvis; Vector: pUT3-Pac (Pharmacia)
with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI-CGAP-Ch2 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma Grade II. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pUT3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
TGATCAGCGT.
TAG_LIB-UI-H-EZ1
TAG_TISSUE-Pac-2-Chondrosarcoma
TAG_SEQ-ATCTAATTAATG"
BASE COUNT      258 a      143 c      125 g      235 t
ORIGIN
Query Match      50.4%; Score 708.4; DB 14; Length 761;
Best Local Similarity 99.3%; Pred. No. 1.4e-74;
Matches 753; Conservative 0; Mismatches 1; Indels 4; Gaps 4;

OY 575 TTTCATGACGATGCGACTGAGATTAATAAAGCAATGGTCCCGAGCAACAGCTGGTGG 634

```

```

Db      754 TTTTCAGAGAT-GGAGCTGAGATTAAGGCG-ATTGGTCCCGGAGCAACAGCTGGT-G 698
OY      635 TECTGAGAGTCAATGTTAAATTCAGAGCACTCCAGTCAAGAGTCAAGTGGAGGTG 694
Db      697 TECTGAGAGTCAATGTTAAATTCAGAGCACTCCAGTCAAGAGTCAAGTGGAGGTG 638
OY      695 CTGCTAAATTTTCCCTCCATCCCTTTTCAGAGCAATGAAATTCGAATCCCACT 754
Db      637 CTGCTAAATTTTCCCTCCATCCCTTTTCAGAGCAATGAA-TTGCAATCTGAACCACT 579
OY      755 GAAAAACAAATTTGCTGAAATTTGAGTATGAGTCCAGTCAAGATTTCTTACCG 814
Db      578 GAAAAACAAATTTGCTGAAATTTGAGTATGAGTCCAGTCAAGATTTCTTACCG 519
OY      815 TCTCCACAAAGTCAAGATTTGAAATGCTCAATCTGCTTTTCTTTTCTTCTTCC 874
Db      518 TCTCCACAAAGTCAAGATTTGAAATGCTCAATCTGCTTTTCTTTTCTTCTTCC 459
OY      875 TCAAGACAGTCAATCTGCTTTCAGAGCTGTTTAAACCTTTGCTGCTGTTTAA 934
Db      458 TCAAGACAGTCAATCTGCTTTCAGAGCTGTTTAAACCTTTGCTGCTGTTTAA 399
OY      935 TATGCTGATTCCTGCTTTCCTGCTTTCAGAGCTGTTTCCGCTGCTGCTGAG 994
Db      398 TATGCTGATTCCTGCTTTCCTGCTTTCAGAGCTGTTTCCGCTGCTGCTGAG 339
OY      995 TATATTTTCTTTGATTTATTTATGAGTTCAGATTTAGTTCAGTTCCTGAG 1054
Db      338 TATATTTTCTTTGATTTATTTATGAGTTCAGATTTAGTTCAGTTCCTGAG 279
OY      1055 TCAATCTCAAGCTTTTGTATCAACAGCAAGAGCTGCTGCTGCTTCCATG 1114
Db      278 TCAATCTCAAGCTTTTGTATCAACAGCAAGAGCTGCTGCTGCTTCCATG 219
OY      1115 AGTTAGTGAATGTTATTTATGAGTTCAGATTTAGTTCAGTTCCTGAGT 1174
Db      218 AGTTAGTGAATGTTATTTATGAGTTCAGATTTAGTTCAGTTCCTGAGT 159
OY      1175 ATGGAAGATTTACATCTGATTTATTTATGATTTCTTCTTCTTCTTCTT 1234
Db      158 ATGGAAGATTTACATCTGATTTATTTATGATTTCTTCTTCTTCTTCTT 99
OY      1235 AGAGATTTGTAATTTGTCAGACTATGTAACAAACAGTGAAGATTTAT 1294
Db      98 AGAGATTTGTAATTTGTCAGACTATGTAACAAACAGTGAAGATTTAT 39
OY      1295 ATATTTCTTATTTGGAATTAAGTAAATTAAGTAAATTAAGTAAAT 1332
Db      38 ATATTTCTTATTTGGAATTAAGTAAATTAAGTAAATTAAGTAAAT 1

```

```

RESULT 7
AL530265 945 bp mRNA linear EST 13-FEB-2001
LOCUS AL530265 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD009YH19 5
DEFINITION prime, mRNA sequence.
ACCESSION AL530265
VERSION AL530265.1 GI:12793758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 945)
AUTHORS Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers

```

```

source 1. 945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1db="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by life technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@life.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 280 a 186 c 224 g 246 t 9 others
ORIGIN

```

```

Query Match 50.4%; Score 707.8; DB 9; Length 945;
Best Local Similarity 88.4%; Pred. No. 1.8e-74;
Matches 835; Conservative 5; Mismatches 7; Indels 98; Gaps 3;

```

```

OY      60 CGTTGCTCTCCGGAACAGCTATCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 119
Db      1 CGTTGCTCTCCGGAACAGCTATCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 60
OY      120 AGTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 179
Db      61 AGTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
OY      180 ATGTCAGAGTAAATCCGAAATATGATTTATTTCAATGTTCTGATTTGGCAGTCA 239
Db      121 ATGTCAGAGTAAATCCGAAATATGATTTATTTCAATGTTCTGATTTGGCAGTCA 180
OY      240 GGGGTTGGAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 299
Db      181 GGGGTTGGAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 240
OY      300 ATCAGCACAAATGGTGTGATTTCAAAATTAAGATTTAGATTTAGAGGCGGAAACAT 359
Db      241 ATCAGCACAAATGGTGTGATTTCAAAATTAAGATTTAGATTTAGAGGCGGAAACAT 300
OY      360 AGCTTCAATAA----- 371
Db      301 AGCTTCAATAATGGGACACAGCGCAGGAAAGATTTGCAACATGACCTCCAGTTAT 360
OY      372 -----GAGTCTTCAAT 383
Db      361 TACAGAGAGCCCATGCGATCATGTTGTATGATGTGACAGATCAGAGAGCTTCAAT 420
OY      384 AATGTTAAACAGTGGCTGCGAGGAATAGATCGTTTGGCAGGAAATGTAACAAATG 443
Db      421 AATGTTAAACAGTGGCTGCGAGGAATAGATCGTTTGGCAGGAAATGTAACAAATG 480
OY      444 TTGTTAGGGAACAAATGATCTGACCAAGAAAGATGATGATACACAACAGCGAAG 503
Db      481 TTGTTAGGGAACAAATGATCTGACCAAGAAAGATGATGATACACAACAGCGAAG 540
OY      504 GAATTTGCTGATTCCTTGGAAATTCGTTTGGAAACAGTGTAGAAATGCAAGAT 563
Db      541 GAATTTGCTGATTCCTTGGAAATTCGTTTGGAAACAGTGTAGAAATGCAAGAT 600
OY      564 GAGAAACAGTCTTATATACAGATGCGACCTGAGATTAAGAAAGCGGATCCGAGACA 623
Db      601 GTAGAACAGTCTTATATACAGATGCGACCTGAGATTAAGAAAGCGGATCCGAGACA 659
OY      624 ACAGCTGCTGCTGAGAGAGTCAATGTTAAATTCAGAGACATCCAGTCAAGAGATCA 683
Db      660 ACAGCTGCTGCTGAGAGAGTCAATGTTAAATTCAGAGACATCCAGTCAAGAGATCA 719

```

QY 684 -GGTGGAGTTCGTCGTAATAATTTGGCCATCCCTTTCTCAGCAAGTAATTTGCAAT 742  
|||||  
Db 720 GGGTGGAGTTCGTCGTAATAATTTGGCCATCCCTTTCTCAGCAAGTAATTTGCAAT 779  
QY 743 CTGAACCCAGTGAATAAATAAATTTGCGAATTTGCTATGATGAGTGCACATCAAC 802  
|||||  
Db 780 CTGAACCCAGTGAATAAATAAATTTGCGAATTTGCTATGATGAGTGCACATCAAC 839  
QY 803 AGATTCTTACCGTCTCCACAAAGTGCAGATTTGAATGTCAATATGCTACTTTT 862  
|||||  
Db 840 AGATTCTTACCGTCTCCACAAAGTGCAGATTTGAATGTCAATATGCTACTTTT 899  
QY 863 TATTCCTTGACTGACGACAGAGCTACTTCAATTTTCAGAACTGTTT 907  
|||||  
Db 900 TATTCCTTGACTGACGACAGAGCTACTTCAATTTTCAGAACTGTTT 944

RESULT 8  
BM452262 1105 bp mRNA linear EST 05-FEB-2002  
LOCUS BM452262  
DEFINITION AGENCOURT\_6386191 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:552635  
5', mRNA sequence.  
ACCESSION BM452262.1 GI:18501302  
VERSION BM452262.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1105)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph. D.  
Email: cga@bbs-femail.nih.gov  
Tissue Procurement: ATCC/DOCTO/DRP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLNL2200 row: 3 column: 12  
High quality sequence stop: 733.  
Location/Qualifiers  
FEATURES  
Source 1..1105  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:552635"  
/clone\_id="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; site\_1: NotI;  
site\_2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
BASE COUNT 320 a 216 c 279 g 282 t 8 others  
ORIGIN

Query Match 50.1%; Score 704.4; DB 13; Length 1105;  
Best Local Similarity 86.0%; Pred. No. 4.1e-74;  
Matches 899; Conservative 0; Mismatches 43; Indels 103; Gaps 7;

Db 122 AGTACGGGGGGGGGCTGGCGGACAGAGGGCGCGGCGGCGGAGCTGCAGTCAC 181  
QY 180 ATGTCCAGATTAATCCCGAATATGATTTATTATTCAGATTACTTGATGGCGATCA 239  
Db 182 ATGTCCAGATTAATCCCGAATATGATTTATTATTCAGATTACTTGATGGCGATCA 241  
QY 240 GGGGTGGAAAGTCTTGGCTCTCTCTTAAAGTTTGCAGATGATATATACAGAAAGCTAC 299  
Db 242 GGGGTGGAAAGTCTTGGCTCTCTCTTAAAGTTTGCAGATGATATATACAGAAAGCTAC 301  
QY 300 ATCAGCAATATGATGATTTCAAAATAGAACTATAGATTAGACGGGAAAGCAATC 359  
Db 302 ATCAGCAATATGATGATTTCAAAATAGAACTATAGATTAGACGGGAAAGCAATC 361  
QY 360 AAGGTTCAATAA----- 371  
Db 362 AAGGTTCAATAATGGGACAGACGCGGAGGAAAGATTTCGAACATCACCCTCAATAT 421  
QY 372 -----GAGTCCCTTCAT 383  
Db 422 TACAGAGAGCCCATGCGATCATAGTTGTATGATGTGACAGATCATGAGATCCTTCAT 481  
QY 384 AATGTTAAACAGTGGCTGCAGAGAAATAGATGATTCAGGAAATGTCAACAAATG 443  
Db 482 AATGTTAAACAGTGGCTGCAGAGAAATAGATGATTCAGGAAATGTCAACAAATG 541  
QY 444 TTGGTGGGAAACAAATGATGATGACCAAGAAAGATGATGATACCAACAGGAG 503  
Db 542 TTGGTGGGAAACAAATGATGATGACCAAGAAAGATGATGATACCAACAGGAG 601  
QY 504 GAATTTGCGATTCCTTGGAAATTCGTTTGGAAACGAGCTGTAAGAAAGCAAGAT 563  
Db 602 GAATTTGCGATTCCTTGGAAATTCGTTTGGAAACGAGCTGTAAGAAAGCAAGAT 661  
QY 564 GTGAACACTCTTTCATGACATGAGAGTGAATTTAAAGCGAATGGTCCCGAGACA 623  
Db 662 GTGAACACTCTTTCATGACATGAGAGTGAATTTAAAGCGAATGGTCCCGAGACA 721  
QY 624 ACAGCTGGTGGTCTGAGAAAGTCAATGTTAAATTCAGACATCTCAGTCAAGACATCA 683  
Db 722 ACAGCTGGTGGTCTGAGAAAGTCAATGTTAAATTCAGACATCTCAGTCAAGACATCA 780  
QY 684 -GGGAGAGTTCGCTGTAATAATTTGGCTCCATCCTTTCTCAGCAAGTAATTTGCAAT 742  
Db 781 NGGTGAGAGTTCGCTGTAATAATTTGGCTCCATCCTTTCTCAGCAAGTAATTTGCAAT 840  
QY 743 CTGAACCCAGTGAATAAATAAATTTGGCTGATTTGATGATGATGATGATGATGATGAT 802  
Db 841 TGAACCCAGTGAATAAATAAATAAATTTGGCTGATTTGATGATGATGATGATGATGAT 899  
QY 803 AGATTCTTACCGTCTCCACAAAGTGCAGATTTGAATGTCAATATGCTACTTTT 860  
Db 900 AGATTCTTACCGTCTCCACAAAGTGCAGATTTGAATGTCAATATGCTACTTTT 959  
QY 861 TTTAATTCCTTGACGACAGACAGCTACTTCAATTTTCAGACGTTTAAAC-TTTGTG 919  
Db 960 TTTAATTCCTTGACGACAGACAGCTACTTCAATTTTCAGACGTTTAAACCTTTTGTG 1019  
QY 920 TCGTGGTTTATTAATATGTTGT 944  
Db 1020 TCGTGGTTTATTAATATGTTGT 1044

RESULT 9  
BM448090/c 725 bp mRNA linear EST 28-MAY-2002  
LOCUS BM448090  
DEFINITION UT-H-FU1-bai-k-14-0-UT.51 NCI-CGAP\_Ct1 Homo sapiens cDNA clone  
UT-H-FU1-bai-k-14-0-UT 3', mRNA sequence.  
ACCESSION BM448090  
VERSION BM448090.1 GI:21251202  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 725)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. Jose Mercuende  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
Seq primer: MJ3 FORWARD  
POLYA-Yes.

## FEATURES

Source  
Location/Qualifiers

1..725  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-H-E01-Dal-K-14-0-UI"  
/clone.lib="NCI.CGAP.Ctl1"  
/tissue.type="Osteoarthritic Cartilage"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Knee; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site.1: Ecor I; Site.2: Not I; NCI.CGAP.Ctl1 is a normalized cDNA library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TCATCAGCT.  
TAG\_LIB=UI-H-E01  
TAG\_TISSUE="Osteoarthritic cartilage"  
TAG\_SEQ="TCATCAGCT"  
BASE COUNT 247 a 135 c 120 g 221 t 2 others  
ORIGIN

Query Match 50.1%; Score 703.6; DB 14; Length 725;  
Best Local Similarity 99.0%; Pred. No. 6.8e-74;  
Matches 717; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 609 ATGGGTCCCGAGCAACAGCTGGTGGTCTGAGAAAGTCAATGTTAAATTCAGAGCACT 668  
DB 723 AGGGGTCCCGAGCAACAGCTGTGNNGTGAGAGTCAATGTTAAATTCAGAGCACT 664  
QY 669 CCAGCAAGCACTGAGAGTGGTGGTCTGAGAAAGTCAATGTTAAATTCAGAGCACT 728  
DB 663 CCAGCAAGCACTGAGAGTGGTGGTCTGAGAAAGTCAATGTTAAATTCAGAGCACT 605  
QY 729 AATGATTTGCAATCTGAACCAAGTGAACCAATGTTAAATTCAGAGCACT 788  
DB 604 AATGATTTGCAATCTGAACCAAGTGAACCAATGTTAAATTCAGAGCACT 545  
QY 789 AGCTGACTACACAGATTTTACCGCTCTCCACAAAGTCAAGATTTGTAATGGTCAAT 848  
DB 544 AGCTGACTACACAGATTTTACCGCTCTCCACAAAGTCAAGATTTGTAATGGTCAAT 485  
QY 849 ACTGACTTTTATATCCCTTGACTCAAGACAGTCAATCTGATTTTCAAGAGCTTTT 908  
DB 484 ACTGACTTTTATATCCCTTGACTCAAGACAGTCAATCTGATTTTCAAGAGCTTTT 425  
QY 909 AAACCTTTGTGCTGTTTATTAATATGTTGTTAAATCCCTGTTTCTGATATCC 968  
DB 424 AAACCTTTGTGCTGTTTATTAATATGTTGTTAAATCCCTGTTTCTGATATCC 365

QY 969 AGACGTGTTCCCGTGGTGGTGAATATATTTGTTGATGTTATATGGCATGTT 1028  
DB 364 AGACGTGTTCCCGTGGTGGTGAATATATTTGTTGATGTTATATGGCATGTT 305  
QY 1029 AGATGTCAAGTGTAGTCTTCTGAGAGTGAAGTTCAGCCATTTTGTATCAACAGCAAG 1088  
DB 304 AGATGTCAAGTGTAGTCTTCTGAGAGTGAAGTTCAGCCATTTTGTATCAACAGCAAG 245  
QY 1089 CAGTGTGCTGACTTCCATGATGATTAAGTGTAGATGATATGTAATGATGAT 1148  
DB 244 CAGTGTGCTGACTTCCATGATGATTAAGTGTAGATGATATGTAATGATGAT 185  
QY 1149 GCTAGTCTTCTCTTGTAGAGTATTAATGAAAGTTCACATCTGATTTATGTTCT 1208  
DB 184 GCTAGTCTTCTCTTGTAGAGTATTAATGAAAGTTCACATCTGATTTATGTTCT 125  
QY 1209 TCATCTCTGCAATATATTTGCTGCGAGATATGTAATTTGTCACACTATGTAAC 1268  
DB 124 TCATCTCTGCAATATATTTGCTGCGAGATATGTAATTTGTCACACTATGTAAC 65  
QY 1269 AAAACACTGAAGATATGTTTATTAATATTTGATTTTGGAGTAAAAA 1328  
DB 64 AAAACACTGAAGATATGTTTATTAATATTTGATTTTGGAGTAAAAA 5  
QY 1329 AAAA 1332  
DB 4 AAAA 1

RESULT 10  
BI913092  
LOCUS  
DEFINITION  
603179695F1 NIH\_MGC\_121 Homo sapiens cDNA clone IMAGE:5244094.5', mRNA sequence.  
BI913092  
ACCESSION  
BI913092.1 GI:16177393  
VERSION  
BI913092.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 796)  
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM11615 row: m column: 23  
High quality sequence stop: 784.

## FEATURES

Source  
Location/Qualifiers

1..796  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5244094"  
/clone.lib="NIH\_MGC\_121"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: EcorV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.7 kb. Insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note:

BASE COUNT 235 a 140 c 168 g 253 t  
 ORIGIN this is a NIH-MGC library \*

Query Match 50.0%; Score 702; DB 13; Length 796;  
 Best Local Similarity 95.7%; Pred. No. 9, 8e-74;  
 Matches 734; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

485 AGACTACACACACGAGAGAAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCG 544  
 1 AGACTACACACACGAGAGAA-TTGCTGATTCCTTGGAAATTCCTTTTGGAAACCG 57  
 Db  
 545 TGTCTAGATGACAGCAATGTAGAAAGCTCTTCATGACAGTGCAGCTGACATTTAAAA 604  
 58 TGTCTAGATGACAGCAATGTAGAAAGCTCTTCATGACAGTGCAGCTGACATTTAAAA 117  
 Db  
 605 GCGAATGGTCCCGAGACAGCTGGTGGTCTGAGAAAGTCCAAATGTTAAAAATTCAG 664  
 118 GCGAATGGTCCCGAGACAGCTGGTGGTCTGAGAAAGTCCAAATGTTAAAAATTCAG 177  
 Db  
 665 CACTCCAGTCACAGCAGTGCAGTGGAGTTCGCTAAATTTGCTCCATCCCTTTCTCA 724  
 178 CACTCCAGTCACAGCAGTGCAGTGGAGTTCGCTAAATTTGCTCCATCCCTTTCTCA 237  
 Db  
 725 CAGCAATGATTTGCAATCTGATCCCAAGTGAAG-AAAAGAAATTCGCAATTTGCTG 783  
 238 CAGCAATGATTTGCAATCTGATCCCAAGTGAAG-AAAAGAAATTCGCAATTTGCTG 297  
 Db  
 784 TATGTAGCTCAGCTACACAGATTTCTTACCGTCTCCCAAGAAAGTGCAGATTTGTAATG 843  
 298 TATGTAGCTCAGCTACACAGATTTCTTACCGTCTCCCAAGAAAGTGCAGATTTGTAATG 357  
 Db  
 844 TCATATAGTACTTTTATTTTATTCCTGACACAGCAGTACCTTCAATTTTCAGAACT 903  
 358 TCATATAGTACTTTTATTTTATTCCTGACACAGCAGTACCTTCAATTTTCAGAACT 417  
 Db  
 904 GTTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 963  
 418 GTTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 477  
 Db  
 964 AATACCAACCTTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1023  
 478 AATACCAACCTTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537  
 Db  
 1024 TGTCTAGATGCTCAGTTCAGTTCCTGAGAAAGTTCAGCCATTTTGTATCAACAGC 1083  
 538 TGTCTAGATGCTCAGTTCAGTTCCTGAGAAAGTTCAGCCATTTTGTATCAACAGC 597  
 Db  
 1084 ACAAGCAGTGTCTGCTCAGTTCCTGAGAAAGTTCAGTTCAGTTCAGTTCAGTTC 1143  
 598 ACAAGCAGTGTCTGCTCAGTTCCTGAGAAAGTTCAGTTCAGTTCAGTTCAGTTC 657  
 Db  
 1144 GATTGCTAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1203  
 658 GATTGCTAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717  
 Db  
 1204 TTTCTCATCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1263  
 718 TTTCTCATCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 777  
 Db  
 1264 GTACAAA 1271  
 778 GTACAAA 785  
 Db

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 760)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subcloning: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccray@uiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 FORWARD  
 POLYA=yes.

FEATURES  
 source  
 location/Qualifiers  
 1..760  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UT-CF-BC1-ab1-a-17-0-UT"  
 /clone\_1lb="UT-CF-BC1"  
 /tissue\_type="lung"  
 /dev\_stage="Adult and Fetal"  
 /lab\_instrument="Ph10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoR I; Site\_2: Not I;  
 UT-CF-BC1 is a normalized cDNA library containing the  
 following tissue(s): Normal lung from adult and from fetal  
 day 64, day 87, week 19 and week 42. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dr)18 tail. The sequence tag for this library is  
 AAGTCTTAC.  
 TAG\_LIB-UT-CF-BC1  
 TAG\_TISSUE=Normal lung Epithelial Cells Tissue nos 369-371  
 and 380-383  
 TAG\_SEQ=AAGTCTTAC"

BASE COUNT 255 a 142 c 125 g 236 t 2 others  
 ORIGIN

Query Match 49.3%; Score 692.8; DB 14; Length 760;  
 Best Local Similarity 99.1%; Pred. No. 1, 2e-72;  
 Matches 738; Conservative 0; Mismatches 3; Indels 4; Gaps 4;

588 GCAGCGAGATTAAAGGATGGTCCCGAGCAACAGCTGGTGGTGTGAGAGTCC 647  
 741 GCAGCGAGATTAAAGGATGGTCCCGAGCAACAGCTGGTGGTGGTGTGAGAGTCC 684  
 Db  
 648 AATGTTAAATTTGAGCACTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 707  
 683 AATGTTAAATTTGAGCACTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 766  
 Db  
 708 CCGCACTCTTTTTCAGCAGCAAGTATTTGCAATGTGACCAAGTGAAGAAAAACAAT 767  
 625 CCGCACTCTTTTTCAGCAGCAAGTATTTGCAATGTGACCAAGTGAAGAAAAACAAT 566  
 Db  
 768 TGCTGAAATTTGATGATGCTGACACAGATTTTACGCTCTCCCAAGGT 827  
 Db

RESULT 11  
 BM971301/c 760 bp mRNA linear EST 21-MAR-2002  
 LOCUS UI-CF-BC1-ab1-a-17-0-UT.82 UT-CF-BC1 Homo sapiens cDNA clone  
 DEFINITION UI-CF-BC1-ab1-a-17-0-UT 3', mRNA sequence.  
 ACCESSION BM971301  
 VERSION BM971301.1 GI:19588888  
 KEYWORDS EST.

```

|||||
565 TCCGGAATGACGTATAGTACGACATACAGAGATCTACCGCTCCACCAAGAT 506
828 CAGAGATGTAATGGTCAATAGTCTTTTATTCCTGACCTCAAGACACTAA 887
505 CAGAGATGTAATGGTCAATAGTCTTTTATTCCTGACCTCAAGACACTAA 446
888 CTTCATTTTCAGAACTGTTTAACTGTTGCTGCTGTTTAAATATATGCTGTAAT 947
445 CTTCATTTTCAGAACTGTTTAACTGTTGCTGCTGTTTAAATATATGCTGTAAT 386
948 CCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007
385 CCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 326
1008 GATGTTATATGCGCATGTTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1067
325 GATGTTATATGCGCATGTTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 266
1068 TTTTGTATCAAAACAGACAGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1127
265 TTTTGTATCAAAACAGACAGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 206
1128 GTTATATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1187
205 GTTATATGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 146
1188 ACTATCTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1247
145 ACTATCTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 86
1248 ATTGTGTGACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1307
85 ATTGTGTGACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 26
1308 TGGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1332
25 TGGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1

```

**RESULT 12**  
**BM919860**  
**LOCUS** AGENCOURT\_6708189 NIH\_MGC\_120 Homo sapiens cDNA clone IMAGE:5749635  
**DEFINITION** 5', mRNA sequence.  
**ACCESSION** BM919860  
**VERSION** BM919860.1 GI:19370239  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
**REFERENCE** NIH-MGC http://mgi.nci.nih.gov/  
**AUTHORS** 1 (bases 1 to 976)  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-riemail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM12779 row: f column: 04  
 High quality sequence stop: 757.  
**FEATURES**  
 source  
 1..976  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5749635"  
 /clone\_1ib="NIH\_MGC\_120"

```

/1ab_host="DH10B"
/Note="Organ: pooled pancreas and spleen; Vector:
pcmv-SPOrt6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(boovr site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH-MGC library."
BASE COUNT      285 a      196 c      246 g      248 t      1 others
ORIGIN
Query Match      49.2%; Score 691.8; DB 14; Length 976;
Best Local Similarity 87.6%; Pred. No. 1.3e-72;
Matches 832; Conservative 0; Mismatches 18; Indels 100; Gaps 3;
16 GCGGCGGCTGCTGATGTTGTTCTAGGAGCAGTAGGGGAAGAGCTTGTCTCCGGA 75
|||||
13 GCGGCGGCTGCTGATGTTGTTCTAGGAGCAGTAGGGGAAGAGCTTGTCTCCGGA 72
|||||
76 AGAGCTATCTGATTCCTTCTTGTGATTAACCGTGGCGGAGAGTCAAGGCGCGCT 135
|||||
73 AGAGCTATCTGATTCCTTCTTGTGATTAACCGTGGCGGAGAGTCAAGGCGCGCT 132
|||||
136 GCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 195
|||||
133 GCGGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 192
|||||
196 CCGATATGATTTATTTCAAGTACTTCTGATTTGGGAGCTCAGGGGTTGGAAGTCTT 255
|||||
193 CCGAATATGATTTATTTCAAGTACTTCTGATTTGGGAGCTCAGGGGTTGGAAGTCTT 252
|||||
256 GCCTTCTTGTGTTGTTGAGATGATATATACAAAGTATATACAAATTTGGG 315
|||||
253 GCCTTCTTGTGTTGTTGAGATGATATATACAAAGTATATACAAATTTGGG 312
|||||
316 TGGATTTCAAAATTAAGAACTATAGAGTTAGAGGGAACAAATCAAGCTTCAATA 371
|||||
313 TGGATTTCAAAATTAAGAACTATAGAGTTAGAGGGAACAAATCAAGCTTCAATA 372
|||||
372 ----- 371
373 ACACAGCAGGCGCAGAAAGATTTGCAACATCACTGATTTATTAACAGAGGCCATG 432
372 ----- GAGTCTTCAATATGTTAAAGTGGC 399
433 GCATCATGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 492
|||||
400 TGCAGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 459
|||||
493 TGCAGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 552
|||||
460 GTGATCTGACCAAAAGAAAGTAGATATACCAACAGCAGGAAGATTTGCTGATGCC 519
|||||
553 GTGATCTGACCAAAAGAAAGTAGATATACCAACAGCAGGAAGATTTGCTGATGCC 612
|||||
520 TTGGAATTCCTTTTGGAAACAGTGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 579
|||||
613 TTGGAATTCCTTTTGGAAACAGTGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 672
|||||
580 TGACGATGCGACGTAGATTAAGAAAGCAAGTGGTCCGAGCAACAGCTGGTGGCTG 639
|||||
673 TGACGATGCGACGTAGATTAAGAAAGCAAGTGGTCCGAGCAACAGCTGGTGGCTG 732
|||||
640 AGAAGTCCAAATGTTAAATTCAGAGCAGTCCAGTCAAGAGTGGAGTGGTGGCTG 699
|||||
733 AGAAGTCCAAATGTTAAATTCAGAGCAGTCCAGTCAAGAGTGGAGTGGTGGCTG 792
|||||
700 AAAATTTGCTCCTATCCTTTCTCAGAGCAAGATTTGCAATGCAATCCCAAGTAA 759
|||||
793 AAAATTTGCTCCTATCCTTTCTCAGAGCAAGATTTGCAATGCAATCCCAAGTAA 852

```





```

/tissue_type="lens"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies (T1 phage resistant))"
/Note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-DMO is a cDNA library containing the following
tissue(s): lens. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CGATTGCGGA. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT      195 a      122 c      142 g      230 t
ORIGIN

```

```

Query Match      49.0%; Score 689; DB 14; Length 689;
Best Local Similarity 100.0%; Pred. No. 3.6e-72;
Matches 689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 503 GGAATTTGCGATTCCTTGAATTCCTTTTGGAAACAGTGTAAAGTCAACGAA 562
    |||
DB 1  GGAATTTGCTATTCCTTGAATTCCTTTTGGAAACAGTGTAAAGTCAACGAA 60
    |||
OY 563 TGTGAACAGTCTTTCATGAGATGGCAGTGTAAAGTCAACGAAAGGTCCTCCGAGC 622
    |||
DB 61  TGTGAACAGTCTTTCATGAGATGGCAGTGTAAAGTCAACGAAAGGTCCTCCGAGC 120
    |||
OY 623 AACAGCTGTGTGTCTGAGAGTCCATGTAAATTCAGAGCACTCCAGTCAAGCAGT 682
    |||
DB 121 AACAGCTGTGTGTCTGAGAGTCCATGTAAATTCAGAGCACTCCAGTCAAGCAGT 180
    |||
OY 683 AGGTGAGAGTGTGTCTGAGAGTCCATGTAAATTCAGAGCACTCCAGTCAAGT 742
    |||
DB 181 AGGTGAGAGTGTGTCTGAGAGTCCATGTAAATTCAGAGCACTCCAGTCAAGT 240
    |||
OY 743 CTGAACCCAGTGTAAAGTCAACGAAATTCCTTTCATGAGATGGCAGTGTAAAGT 802
    |||
DB 241 CTGAACCCAGTGTAAAGTCAACGAAATTCCTTTCATGAGATGGCAGTGTAAAGT 300
    |||
OY 803 AGATTTCTTACCGTCTCCAGAGTGTCAAGATTTGAATGTGCAATGCACTTTT 862
    |||
DB 301 AGATTTCTTACCGTCTCCAGAGTGTCAAGATTTGAATGTGCAATGCACTTTT 360
    |||
OY 863 TATTCCTTCTGACTCAAGAGTGTCAAGATTTGAATGTGCAATGCACTTTT 922
    |||
DB 361 TATTCCTTCTGACTCAAGAGTGTCAAGATTTGAATGTGCAATGCACTTTT 420
    |||
OY 923 TGGTTTAAATTAATGTGTAAATTCCTTTCATGAGATGGCAGTGTAAAGT 982
    |||
DB 421 TGGTTTAAATTAATGTGTAAATTCCTTTCATGAGATGGCAGTGTAAAGT 480
    |||
OY 983 GGTGTTGAGAAATATTTTGTGTGAATTTGAATGTGCAATGCACTTTT 1042
    |||
DB 481 GGTGTTGAGAAATATTTTGTGTGAATTTGAATGTGCAATGCACTTTT 540
    |||
OY 1043 GCTCTTGAAGATTAATGTGTGAATTTGAATGTGCAATGCACTTTT 1102
    |||
DB 541 GCTCTTGAAGATTAATGTGTGAATTTGAATGTGCAATGCACTTTT 600
    |||
OY 1103 TTTCATGATTAATGTGTGAATTTGAATGTGCAATGCACTTTT 1162
    |||
DB 601 TTTCATGATTAATGTGTGAATTTGAATGTGCAATGCACTTTT 660
    |||
OY 1163 GTAGAGTTATTAATGTGAAGATTTACACTA 1191
    |||
DB 661 GTAGAGTTATTAATGTGAAGATTTACACTA 689
    |||

```

```

RESULT 15
BI092003      737 bp      mRNA      linear      EST 20-JUN-2001
LOCUS        60285630221 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4997688 5',
DEFINITION   mRNA sequence.
ACCESSION    BI092003
VERSION      BI092003.1 GI:14510333
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 737)
AUTHORS      NIH-MGC http://mhc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-remail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLNL11025 row: c column: 01
              High quality sequence stop: 727.

```

## FEATURES

```

source
Location/Qualifiers
1..737
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4997688"
/clone_id="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/Note="Organ: cervix; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

```

```

BASE COUNT      214 a      135 c      147 g      241 t
ORIGIN

```

```

Query Match      49.0%; Score 688.2; DB 13; Length 737;
Best Local Similarity 97.8%; Pred. No. 4.3e-72;
Matches 719; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

```

```

OY 550 AGAATGCAACGAATGTAGAACAGTCTTTCATGAGATGGCAGTGTAAAGTCAACGAA 609
    |||
DB 1  AGAATGCAACGAATGTAGAACAGTCTTTCATGAGATGGCAGTGTAAAGTCAACGAA 60
    |||
OY 610 TGGTCCCGAGCAACAGCTGGTGTGCTGAGAGTCCATGTTAAATTCAGAGCACTC 669
    |||
DB 61  TGGTCCCGAGCAACAGCTGGTGTGCTGAGAGTCCATGTTAAATTCAGAGCACTC 120
    |||
OY 670 CAGTCAAGCACTGAGTGTGAGTGTGCTGTTAAATTTGCTCCATCCTTTTCTCAGACA 729
    |||
DB 121 CAGTCAAGCACTGAGTGTGAGTGTGCTGTTAAATTTGCTCCATCCTTTTCTCAGACA 180
    |||
OY 730 ATGAATTTGCAATCGAATCCCAAGTGAAGAAACAAATTCCTGAATGTGCTGTATGTA 789
    |||
DB 181 ATGAATTTGCAATCGAATCCCAAGTGAAGAAACAAATTCCTGAATGTGCTGTATGTA 240
    |||
OY 790 GCTCAGTCAACAGATTTCTTACCGTCTCCCAAGAGTCAAGATTTGAATGTGTCATA 849
    |||
DB 241 GCTCAGTCAACAGATTTCTTACCGTCTCCCAAGAGTCAAGATTTGAATGTGTCATA 300
    |||
OY 850 CTGACTTTTATTTATTCCTTGAATCAAGACAGTCACTTCAATTTTCAGAACTGTTTGA 909
    |||
DB 301 CTGACTTTTATTTATTCCTTGAATCAAGACAGTCACTTCAATTTTCAGAACTGTTTGA 360
    |||
OY 910 AACCTTTGTGCTGTGTTTAAATTAATGTGTATCCTTTGCTTGTCTGATACCA 969
    |||

```

```
Db 361 AACCTTGTGTGTGTTTAAATAATGATGTAACCTTGTCTTCTCGATACA 420
Qy 970 GACTGTTCOCCTGGTGTGTAATATATTTGTTGA--TGTATATGSCATGT 1027
Db 421 GACTGTTCOCCTGGTGTGTAATATATTTGTTGACTGTATATCTTGGCATGT 480
Qy 1028 TAGATGTCAGCTTTAGTCTGTAAGATGCAATTCAGCATTTGTATCAACACACAA 1087
Db 481 TAGATGTCAGCTTTAGTCTGTAAGATGCAATTCAGCATTTGTATCAACACACAA 540
Qy 1088 GCAGTGTCTGTCACTTTCATGCAATAAGTTAGAGATGTTATATGTAAGATCTGATT 1147
Db 541 GCAGTGTCTGTCACTTTCATGCAATAAGTTAGAGATGTTATATGTAAGATCTGATT 600
Qy 1148 TGTAGTCTTCCCTGTAGAGTATAAATGAA--AGATTACCTATCTGATTAAGTTT 1206
Db 601 TGTAGTCTTCCCTGTAGAGTATAAATGAAAGATTAAGTATGATTAAGTTT 660
Qy 1207 CTTCATCTCTGCAATATATTTGTGCTGCAGAAATATGTAATTTGTGCACACTATGTA 1266
Db 661 CTTCATCTCTGCAATATATTTGTGCTGCAGAAATATGTAATTTGTGCACACTATGTA 720
Qy 1267 ACAAACCACTGAG 1281
Db 721 ACAAACCACTGAGC 735
```

Search completed: June 22, 2003, 21:26:05  
Job time : 1910 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2003, 19:45:31 ; Search time 94 Seconds  
(Without alignments)  
4583.839 Million cell updates/sec

Title: US-09-820-003b-1

Perfect score: 1405  
Sequence: 1 aagcgatagctgagtcgagc.....aaaaaaaaaaaaaaaaa 1405

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents.NA:\*  
1: /cgn2.6/pdata1/lna/5A.COMB.seq:\*  
2: /cgn2.6/pdata1/lna/5B.COMB.seq:\*  
3: /cgn2.6/pdata1/lna/6A.COMB.seq:\*  
4: /cgn2.6/pdata1/lna/6B.COMB.seq:\*  
5: /cgn2.6/pdata1/lna/PCNUS.COMB.seq:\*  
6: /cgn2.6/pdata1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195	13.9	925	2	US-08-916-901-4
2	195	13.9	925	2	US-09-154-602-4
3	171	12.2	171	4	US-09-506-729-59
4	166	11.8	166	4	US-09-506-729-58
5	102.6	7.3	2246	4	US-09-363-708-3
6	102.2	7.3	140	1	US-08-628-417-5
7	102.2	7.3	240	1	US-08-628-417-6
8	102	7.3	1872	4	US-09-801-052-1
9	100.8	7.2	1117	4	US-09-247-3738-33
10	100.4	7.1	1166	1	US-08-157-101A-4
11	100.4	7.1	1813	4	US-09-071-224-3
12	100	7.1	1411	4	US-08-964-127-5
13	100	7.1	1411	4	US-09-496-692-5
14	99.8	7.1	1582	3	US-08-545-196B-10
15	99.8	7.1	1582	3	US-08-545-196B-12
16	99.8	7.1	2447	2	US-09-014-969-14
17	99.2	7.1	1474	1	US-08-702-344-26
18	99.2	7.1	1474	1	US-08-821-994-64
19	99.2	7.1	1683	4	US-09-347-803-11
20	99.2	7.1	3375	4	US-09-370-838-151
21	99	7.0	2671	6	5168051-9
22	98.6	7.0	1798	4	US-09-797-906-1
23	98.4	7.0	2806	4	US-09-653-838-9
24	97.8	7.0	2674	4	US-09-817-180-1
25	97.6	6.9	1447	4	US-09-443-041A-27
26	97.4	6.9	578	4	US-09-602-877A-95
27	96.8	6.9	1641	1	US-08-300-903A-8

28	95.6	6.8	1098	3	US-09-248-335-35	Sequence 35, Appl
29	95.6	6.8	2323	4	US-09-148-476-24	Sequence 24, Appl
30	95.2	6.8	1882	4	US-09-370-253-1	Sequence 1, Appl
31	95.2	6.8	2634	4	US-09-463-238-3	Sequence 3, Appl
32	95.2	6.8	6671	1	US-08-280-443-1	Sequence 1, Appl
33	95.2	6.8	6671	1	US-08-457-459-1	Sequence 1, Appl
34	95.2	6.8	6671	1	US-08-555-678-1	Sequence 1, Appl
35	95.2	6.8	6671	5	PCT-US95-02275-1	Sequence 66, Appl
36	94.8	6.7	1864	4	US-09-360-545-66	Sequence 1, Appl
37	94.4	6.7	1454	4	US-09-372-422A-19	Sequence 1, Appl
38	94.2	6.7	1733	3	US-09-073-569-1	Sequence 1, Appl
39	93.8	6.7	1051	4	US-09-245-041-10	Sequence 10, Appl
40	93.8	6.7	2184	4	US-08-955-918C-1	Sequence 1, Appl
41	93.6	6.7	2184	4	US-08-697-766A-1	Sequence 1, Appl
42	93.6	6.7	1493	1	US-08-340-820-24	Sequence 24, Appl
43	93.6	6.7	1493	1	US-08-593-535-24	Sequence 24, Appl
44	93	6.6	2082	2	US-08-785-310A-2	Sequence 2, Appl
45	92.8	6.6	1198	3	US-09-248-335-27	Sequence 27, Appl

#### ALIGNMENTS

RESULT 1  
US-08-916-901-4  
; Sequence 4, Application US/08916901  
; Patent No. 5892012  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley Neil C.  
APPLICANT: Shah, Purya  
TITLE OF INVENTION: RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Inocyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/916,901  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0367 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 925 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LIVERMUT04  
CLONE: 2514506  
US-08-916-901-4  
Query Match 13.9%; Score 195; DB 2; Length 925;  
Best Local Similarity 63.5%; Pred. No. 1e-30; Indels 99; Gaps 2;  
Matches 369; Conservative 0; Mismatches 125;

```

OY      184   CCAGATGAATCCCGAATAATGATTATTTTTCGAAGTACTCTTGATGGGACTCAGGGG    243
Db      61     CGCCCATTAACCCCGCAATATGACTACCTCTTTAAAGCTGTGGATGGGCATAGGGC    120
OY      244   TTGNAAGCTCTGCCCTTCTCTTAGATTGGATGGACAGATGATACATATACAAGAAGCTACATCA    303
Db      121   TGGGCAAGTACGCTGCTGCTCGGCTTGGTATGATGACACGATACAGAGAGGCTACATCA    180
OY      304   GCACATTTGGTGTGATTTTCAAAATTAAGAACTATAGAGTTAGACGGGAAAAACATCAAGC    363
Db      181   GCACCATCGGGGTGAGCTTCAAAGATCCGAACCATCGAGCTGATGCCAAAACATATCAAC    240
OY      364   TTCANAATA-----371
Db      241   TTCAAGATCTGGGACACAGCGGGCCAGAAACGGTTCCGGACATCACTACTCAGCTACTACC    300
OY      372   -----GAGTCCCTTCATTAATG    387
Db      301   GGGGGGCTCATNGGCATCATCTGTTGTATGACGTACTGACCAGAAATCCTTACGCCAACG    360
OY      388   TTAACATGTGGCTGCAGAGAAATAGATCCGTATGCCAGTTAAATGTCAACAAATTGTGG    447
Db      361   TGAACACATGTGGCTGAGGAGATGACCCGTATGCCAGCAAGACGTCATTAACCTCTGG    420
OY      448   TAGGGAACAAATGTATCTATGCACCAAAAGAAAGTAGTAGTACACACAGCAGAAAGAAAT    507
Db      421   TGCGCAACAAGAGCCACCTCACACCACCAAGAAAGTGGTGSACAAACACACAGCCAAAGAGT    480
OY      508   TTGCTGATTCCTTGGAAATTCGTTTTTGGAAACAGAGTCTAAGAAATGCAACGAATGTAG    567
Db      481   TTGCGAGACTCTGTGGGCATCCCTTCTTGGAAACAGAGCCCAAGAAATGCAACCAATGTG    540
OY      568   AACAGCTCTTATGACAGATGAGCAGTGAGATTAATAAAGCAATGGGTCCCAGGACAACAG    627
Db      541   AGCAGGCTTATGACCAATGAGCTCTGAAATCAAAAAGCGAGGGGCTTGAGAGCAGCTT    600
OY      628   CTGGTGTGCTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGACATGACAGTG    687
Db      601   CTGGGGG---GAGAGGGCCCATCTCAAGATGACACGACACCCCTGTAAAGCCGGCTGGCG    657
OY      688   GAGGTGCTGCTA    700
Db      658   GTGGCTGTGGCTA    670

RESULT 2
US-09-154-602-4
: Sequence 4, Application US/09154602
: Patent No. 6300472
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Lal, Preeti
: APPLICANT: Cortley, Neil C.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: RAB PROTEINS
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: IBM Compatible
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: FILING DATE: US/09/154.602
: PRIOR APPLICATION DATA:

```

[illegible]



```

1 ZIP: 21010-5423
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: Patentin Release #1.0, Version #1.25
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/628,417
16
17 FILING DATE:
18
19 CLASSIFICATION: 435
20
21 ATTORNEY/AGENT INFORMATION:
22
23 NAME: BIFONTI, ULYSSES J
24
25 REGISTRATION NUMBER: 39,908
26
27 REFERENCE/DOCKET NUMBER: DAM 398-94
28
29 TELECOMMUNICATION INFORMATION:
30
31 TELEPHONE: 410-671-1158
32
33 TELEFAX: 410-671-2534
34
35 INFORMATION FOR SEQ ID NO: 6:
36
37 SEQUENCE CHARACTERISTICS:
38
39 LENGTH: 240 bases
40
41 TYPE: nucleic acid
42
43 STRANDEDNESS: single
44
45 TOPOLOGY: linear
46
47 MOLECULE TYPE: oligodeoxynucleotide
48
49 HYPOTHEITICAL: NO
50
51 ANTI-SENSE: YES
52
53 US-08-628-417-6
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```





```

TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1411 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...966
US-08-964-127-5

Query Match 7.1%; Score 100; DB 4; Length 1411;
Best Local Similarity 84.8%; Pred. No. 1.3e-11;
Matches 112; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1273 CAACCTGAAGATACCTTTAAATTAATATATCTACTTTATGTGGAAGCTAAAAAAAAAAAAAAAAAAAA 1333
db 1270 CAAATAAAAGACATTGTTGATATAAAGAGCATTTCATAGCGCTGGAAAAA
1329

```

```

Db      1330 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1369
        |||
QY      1393 AAAAAAAAAAAAAA 1404
        |||||
Db      1390 AAAAAAAAAAAAAA 1401

RESULT 13
US-09-496-692-5
; Sequence 5, Application US/09496692
; Patent No. 6313271
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE

```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/496,692  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/964,127  
FILING DATE: 06-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Crews, Ph.D., L. Lee  
REGISTRATION NUMBER: P-43,567  
REFERENCE/DOCKET NUMBER: 07334/038001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1411 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear







Db 241 GGGTTGGAAGTCTTGCCTCTCTCTAGTTTGCAGTATGATATATACAGAAAGCTTCA 300  
 Oy 301 TCAGCACAATTTGGTGTGATTTCAAAATAGAACTATAGAGTTAGACGGGAAACATCA 360  
 Db 301 TCAGCACAATTTGGTGTGATTTCAAAATAGAACTATAGAGTTAGACGGGAAACATCA 360  
 Oy 361 AGCTTCAATAGAGTCTTCAATATATTTAAACATGGCTGCAGAAATAGATCTTATG 420  
 Db 361 AGCTTCAATAGAGTCTTCAATATATTTAAACATGGCTGCAGAAATAGATCTTATG 420  
 Oy 421 CCAGTGAATAATGTCAACAAATTTGTGTAGGAAACAATGTATGTGACCAACAAAG 480  
 Db 421 CCAGTGAATAATGTCAACAAATTTGTGTAGGAAACAATGTATGTGACCAACAAAG 480  
 Oy 481 TAGTAGACTACACAACAGCAGGAAGATTTGCTGATTCCTTGGAAATCCGTTTGGAA 540  
 Db 481 TAGTAGACTACACAACAGCAGGAAGATTTGCTGATTCCTTGGAAATCCGTTTGGAA 540  
 Oy 541 CCAGTGTCTAGAAATGTCAACGAATGTAGAACAGTCTTATGACATGCGACGTGAGTTA 600  
 Db 541 CCAGTGTCTAGAAATGTCAACGAATGTAGAACAGTCTTATGACATGCGACGTGAGTTA 600  
 Oy 601 AAAAGCAATGGGTCCCGGACCAACAGCTGGTGGTGTGAGAAAGTCAATGTTAAATTC 660  
 Db 601 AAAAGCAATGGGTCCCGGACCAACAGCTGGTGGTGTGAGAAAGTCAATGTTAAATTC 660  
 Oy 661 AGAGCACTCCAGTCAAGCAGTCAAGTGGAGTGGTGGTCTGCTAAATTTGCCCTCATCTTT 720  
 Db 661 AGAGCACTCCAGTCAAGCAGTCAAGTGGAGTGGTGGTCTGCTAAATTTGCCCTCATCTTT 720  
 Oy 721 CTCACAGCAATGAAATTTGCAATCTGAAACCCAGTGAATTTGCTGCTGCTGCTGCTGCT 780  
 Db 721 CTCACAGCAATGAAATTTGCAATCTGAAACCCAGTGAATTTGCTGCTGCTGCTGCTGCT 780  
 Oy 781 CTGATAGTAGCTGCACTCAACAGATCTTACCCGCTCCACAAAGTGCAGAGATGTAAA 840  
 Db 781 CTGATAGTAGCTGCACTCAACAGATCTTACCCGCTCCACAAAGTGCAGAGATGTAAA 840  
 Oy 841 TGGTCAATAGTGAATTTTCTTATTCCTTGACTCAAGACAGTCACTTCAATTTTCA 900  
 Db 841 TGGTCAATAGTGAATTTTCTTATTCCTTGACTCAAGACAGTCACTTCAATTTTCA 900  
 Oy 901 ACTGTTTAAACCTTTGTGTGTGTGTATTAATAATGTGTATCTTGTGCTTTC 960  
 Db 901 ACTGTTTAAACCTTTGTGTGTGTGTATTAATAATGTGTATCTTGTGCTTTC 960  
 Oy 961 CTGATACCAAGTCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 Db 961 CTGATACCAAGTCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 Oy 1021 GCATGTTTATGATGCTAGCTTATGCTTGTGAAGATGAAGTCCATTTTGTATCAAC 1080  
 Db 1021 GCATGTTTATGATGCTAGCTTATGCTTGTGAAGATGAAGTCCATTTTGTATCAAC 1080  
 Oy 1081 AGCAACAGCAATGTCTGCTCACTTCCATGCAATGAAGTGTAGAGATTTATATGTA 1140  
 Db 1081 AGCAACAGCAATGTCTGCTCACTTCCATGCAATGAAGTGTAGAGATTTATATGTA 1140  
 Oy 1141 TCTGATTTGTAGTCTTCTTGTAGAGTATTAATGGAAGATTAACATCTGATTA 1200  
 Db 1141 TCTGATTTGTAGTCTTCTTGTAGAGTATTAATGGAAGATTAACATCTGATTA 1200  
 Oy 1201 TAGTTCCTCACTGCTGCAATATATTTGTGCTGCAGAAATATGTAATTTGTGCAC 1260  
 Db 1201 TAGTTCCTCACTGCTGCAATATATTTGTGCTGCAGAAATATGTAATTTGTGCAC 1260  
 Oy 1261 TATGTAAACAAACAGTGAATATGTTAAATATGTAATTTGTGGAAGTGA 1320  
 Db 1261 TATGTAAACAAACAGTGAATATGTTAAATATGTAATTTGTGGAAGTGA 1320  
 Oy 1321 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1380  
 Db 1321 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1380

Oy 1381 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1405  
 Db 1381 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1405  
 RESULT 2  
 US-09-820-003A-3  
 ; Sequence 3, Application US/09820003A  
 ; Patent No. US20020142382A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MERKULOV, Gennady et al.  
 ; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
 ; FILE REFERENCE: CLO01196-  
 ; CURRENT APPLICATION NUMBER: US/09/820,003A  
 ; CURRENT FILING DATE: 2001-03-29  
 ; NUMBER OF SEQ ID NOS: 39  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 46050  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc-feature  
 ; LOCATION: (1)...(46050)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-820-003A-3

Query Match 58.2%; Score 817.2; DB 10; Length 46050;  
 Best Local Similarity 99.6%; Pred. No. 2.4e-130;  
 Matches 819; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Oy 502 AGGAATTTGCTGATCCCTTGGAAATTCCTTTGGAAACCGTGAAGATGCAACGA 561  
 Db 43228 AGGAATTTGCTGATCCCTTGGAAATTCCTTTGGAAACCGTGAAGATGCAACGA 43297  
 Oy 562 ATGTAAGACAGTCTTTCATGACGATGGCAGCTGAGATTTAAAGCGAATGGTCCCGAG 621  
 Db 43298 ATGTAAGACAGTCTTTCATGACGATGGCAGCTGAGATTTAAAGCGAATGGTCCCGAG 43357  
 Oy 622 CAACAGCTGGTGGTCTGAGAGATGCAATGTTAAATTCACAGACATCCAGTCAAGCAGT 681  
 Db 43358 CAACAGCTGGTGGTCTGAGAGATGCAATGTTAAATTCACAGACATCCAGTCAAGCAGT 43417  
 Oy 682 CAGGTGAGGTGGTGGTCTGAAATTTGCTCATCTCTTTCACAGCAATGAATTTGCA 741  
 Db 43418 CAGGTGAGGTGGTGGTCTGAAATTTGCTCATCTCTTTCACAGCAATGAATTTGCA 43477  
 Oy 742 TCTGAACCAAGTGAACCAAAATTCCTGAATGTACTGTATGTAGCTGACTACAA 801  
 Db 43478 TCTGAACCAAGTGAACCAAAATTCCTGAATGTACTGTATGTAGCTGACTACAA 43537  
 Oy 802 CAGATTTCTACCGTCTCCACAAAGTCAAGATTTGAATGTGAATGTGAATGTGACTTTT 861  
 Db 43538 CAGATTTCTACCGTCTCCACAAAGTCAAGATTTGAATGTGAATGTGAATGTGACTTTT 43597  
 Oy 862 TTAATCCCTGACTCAAGACAGTCACTTCAATTTTCAAGTGTGAATGTGAATGTGACTTT 921  
 Db 43598 TTAATCCCTGACTCAAGACAGTCACTTCAATTTTCAAGTGTGAATGTGAATGTGACTTT 43657  
 Oy 922 CTGTTTAAATATATGTGTATATCTTGTCTTCTTCTGATACCAAGTGTTCGCG 981  
 Db 43658 CTGTTTAAATATATGTGTATATCTTGTCTTCTTCTGATACCAAGTGTTCGCG 43717  
 Oy 982 TGTGTTGTTAGAAATATTTTGTGATGTTAATGTGCAATGTGATGATGTCAGGTT 1041  
 Db 43718 TGTGTTGTTAGAAATATTTTGTGATGTTAATGTGCAATGTGATGATGTCAGGTT 43777  
 Oy 1042 AGTCTTGAAGATGAAGTTCAGCAATTTTATCAACACGCAACAGCAGTGTGTCAC 1101  
 Db 43778 AGTCTTGAAGATGAAGTTCAGCAATTTTATCAACACGCAACAGCAGTGTGTCAC 43837









LENGTH: 367  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-040-862-4644

Query Match  
Best Local Similarity 100.0%; Pred. No. 4.9e-34;  
Matches 367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 909 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 968  
DB 1 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 60  
QY 969 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 1028  
DB 61 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 120  
QY 1029 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 1088  
DB 121 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 180  
QY 1089 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 1148  
DB 181 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 240  
QY 1149 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 1208  
DB 241 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 300  
QY 1209 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 1268  
DB 301 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 360  
QY 1269 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 1275  
DB 361 AACCTTTGTCGCTGTTTAAATATGTTAACTCTGTTGCTTCTGATACC 367

RESULT 9  
US-09-918-995-4909  
Sequence 4909, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4909  
LENGTH: 436  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(436)  
OTHER INFORMATION: n - A,T,C or G  
US-09-918-995-4909

Query Match  
Best Local Similarity 24.3%; Score 341.8; DB 9; Length 436;  
Matches 365; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 3 GCGATAGCTGAGT-GCGGCGGCTGCTGATGTTCTAGGAGCAGAGTGGGAGAGC 61  
DB 69 GCGATAGCTGAGT-GCGGCGGCTGCTGATGTTCTAGGAGCAGAGTGGGAGAGC 128  
QY 62 TTGCTCTCCGGAACAGCTATCTATCTTCTTCTGATACCCGCGGAGAGG 121

DB 129 TTGCTCTCCGGAACAGCTATCTATCTTCTTCTGATACCCGCGGAGAG 188  
QY 122 TCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181  
DB 189 TCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 248  
QY 182 GTCACAGATGAATCCGGAATATGATATATATATATATATATATATATAT 241  
DB 249 GTCACAGATGAATCCGGAATATGATATATATATATATATATATATATAT 308  
QY 242 GTCACAGATGAATCCGGAATATGATATATATATATATATATATATATAT 301  
DB 309 GTCACAGATGAATCCGGAATATGATATATATATATATATATATATATAT 368  
QY 302 GTCACAGATGAATCCGGAATATGATATATATATATATATATATATATAT 361  
DB 369 GTCACAGATGAATCCGGAATATGATATATATATATATATATATATATAT 427  
QY 362 GCTCAAT 370  
DB 428 GCTCAAT 436

RESULT 10  
US-09-960-352-5427  
Sequence 5427, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Mengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalaagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511, 006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 5427  
LENGTH: 353  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 23-LIB34-061-Q1-E1-F3  
US-09-960-352-5427

Query Match  
Best Local Similarity 24.0%; Score 337.2; DB 10; Length 353;  
Matches 350; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 946 ATCTTGTGCTTCTCTGATACAGAGCTGTTCCGCTGCTGCTGCTGCTGCT 1005  
DB 1 ATCTTGTGCTTCTCTGATACAGAGCTGTTCCGCTGCTGCTGCTGCTGCT 59  
QY 1006 TTGATGTTATATTTGCGATGTTATGATGATGATGATGATGATGATGATG 1065  
DB 60 TTGATGTTATATTTGCGATGTTATGATGATGATGATGATGATGATGATG 119  
QY 1066 CATTTGTATCAACAGCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1125  
DB 120 CATTTGTATCAACAGCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 179  
QY 1126 ATGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1185  
DB 180 ATGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239  
QY 1186 ATGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1245  
DB 240 ATGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 299  
QY 1246 TAATTTGTTGCACTATGATGATGATGATGATGATGATGATGATGATGAT 1299  
DB 300 TAATTTGTTGCACTATGATGATGATGATGATGATGATGATGATGATGAT 353

RESULT 11  
 US-09-960-352-8911  
 : Sequence 8911, Application US/09960352  
 : Patent No. US20020137139A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Warren, Wesley C.  
 : APPLICANT: Tao, Ningding  
 : APPLICANT: Bylet, John C.  
 : APPLICANT: Mathialagan, Nagappan  
 : TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 : FILE REFERENCE: 16511.006/37-21(10298)C  
 : CURRENT APPLICATION NUMBER: US/09/960,352  
 : CURRENT FILING DATE: 2001-09-24  
 : NUMBER OF SEQ ID NOS: 15112  
 : SEQ ID NO 8911  
 : LENGTH: 427  
 : TYPE: DNA  
 : ORGANISM: Bos taurus  
 : OTHER INFORMATION: Clone ID: 38-LIB34-021-Q1-E1-B6  
 US-09-960-352-8911

Query Match  
 Best Local Similarity 94.7%; Score 314.8; DB 10; Length 427;  
 Pred. No. 4.4e-45;  
 Matches 337; Conservative 0; Mismatches 17; Indels 2; Gaps 1;

QY 18 GCGGCGTCTGATGTGTCTTAGGGACGAGTAGGG--AAGACTTTCCTTCCTCCGGA 75  
 |||  
 DB 1 GCGGCGTCTGATGTGTCTTAGGGACGAGTAGGG--AAGACTTTCCTTCCTCCGGA 60  
 QY 76 ACAGCTTATCTATCTCTTCTTCATTAACCTGCGGCGGAGTCAAGGCGCGGCT 135  
 |||  
 DB 61 ACAGCTTATCTATCTCTTCTTCATTAACCTGCGGCGGAGTCAAGGCGCGGCT 120  
 QY 136 GCGGCGACCAAGGCGCGGCGGCGGCGGCGGAGCTGAGTGCATGTCCAGCATGATC 195  
 |||  
 DB 121 GCAGCAGCAAGGCGCGGCGGCGGCGGCGGAGCTGAGTGCATGTCCAGCATGATC 180  
 QY 196 CCGAATATGATATTTATTTCAAGTACTTCTGATTGGCGACTGAGGCGTTGGAAAGCTT 255  
 |||  
 DB 181 CCGAATATGATATTTATTTCAAGTACTTCTGATTGGCGACTGAGGCGTTGGAAAGCTT 240  
 QY 256 GCGTCTCTTCTTGTGACATGATATATACAGAAAGCTACATCAGCAATTTGGTG 315  
 |||  
 DB 241 GCGTCTCTTCTTGTGACATGATATATACAGAAAGCTACATCAGCAATTTGGTG 300  
 QY 316 TGGATTCAAAATTAAGACTATAGATTAGAGGGAAGAAACATCAAGCTTCAATA 371  
 |||  
 DB 301 TGGATTCAAAATTAAGACTATAGATTAGAGGGAAGAAACATCAAGCTTCAATA 356

RESULT 12  
 US-09-918-995-2704  
 : Sequence 2704, Application US/09918995  
 : Publication No. US20030073623A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Hyseq, Inc.  
 : TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 : FILE REFERENCE: 20411-756  
 : CURRENT APPLICATION NUMBER: US/09/918,995  
 : CURRENT FILING DATE: 2001-07-30  
 : PRIOR APPLICATION NUMBER: US/09/235,076  
 : PRIOR FILING DATE: 1999-01-20  
 : NUMBER OF SEQ ID NOS: 38054  
 : SOFTWARE: FastSeq for Windows Version 3.0  
 : SEQ ID NO 2704  
 : LENGTH: 483  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : NAME/KEY: misc\_feature

LOCATION: (1)...(483)  
 : OTHER INFORMATION: n = A,T,C or G  
 US-09-918-995-2704

Query Match  
 Best Local Similarity 96.9%; Score 314.6; DB 9; Length 483;  
 Pred. No. 5e-45;  
 Matches 317; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 45 CCGAGTAGGGAAGAGCTTGTCTCCCGAAGACGCTATCTCATCTTCTTTCGATT 104  
 |||  
 DB 49 CCGAGTAGGGAAGAGCTTGTCTCCCGAAGACGCTATCTCATCTTCTTTCGATT 108  
 QY 105 ACCCGTGGCGGAGAGTAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 164  
 |||  
 DB 109 ACCCGTGGCGGAGAGTAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 168  
 QY 165 GCGAGCTGAGTGCATGTCCAGCATGATGATGATGATGATGATGATGATGATGAT 224  
 |||  
 DB 169 GCGAGCTGAGTGCATGTCCAGCATGATGATGATGATGATGATGATGATGATGAT 228  
 QY 225 CTGATTGGCGAGTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 284  
 |||  
 DB 229 CTGATTGGCGAGTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 288  
 QY 285 TATACAGAAAGCTTCAATCAGCAGCATTTGGTGGATTCAAAATTAAGACTATGATTA 344  
 |||  
 DB 289 TATACAGAAAGCTTCAATCAGCAGCATTTGGTGGATTCAAAATTAAGACTATGATTA 348  
 QY 345 GAGCGAAACATCAAGCTTCAATA 371  
 |||  
 DB 349 GAGCGAAACATCAAGCTTCAATA 375

RESULT 13  
 US-09-918-995-34631  
 : Sequence 34631, Application US/09918995  
 : Publication No. US20030073623A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Hyseq, Inc.  
 : TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 : FILE REFERENCE: 20411-756  
 : CURRENT APPLICATION NUMBER: US/09/918,995  
 : CURRENT FILING DATE: 2001-07-30  
 : PRIOR APPLICATION NUMBER: US/09/235,076  
 : PRIOR FILING DATE: 1999-01-20  
 : NUMBER OF SEQ ID NOS: 38054  
 : SOFTWARE: FastSeq for Windows Version 3.0  
 : SEQ ID NO 34631  
 : LENGTH: 459  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : NAME/KEY: misc\_feature  
 : LOCATION: (1)...(459)  
 : OTHER INFORMATION: n = A,T,C or G  
 US-09-918-995-34631

Query Match  
 Best Local Similarity 91.4%; Score 280.4; DB 9; Length 459;  
 Pred. No. 3.5e-39;  
 Matches 296; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 48 AGTAGGGGAAGAGCTTGTCTCCCGAAGACGCTATCTCATCTTCTTTCGATTAC 107  
 |||  
 DB 20 AGTAGGGGAAGAGCTTGTCTCCCGAAGACGCTATCTCATCTTCTTTCGATTAC 79  
 QY 108 CGTGGCGGAGAGTAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 167  
 |||  
 DB 80 CGTGGCGGAGAGTAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 139  
 QY 168 AGCTGAGTGCATGTCCAGCATGATGATGATGATGATGATGATGATGATGATGAT 227  
 |||  
 DB 140 AGCTGAGTGCATGTCCAGCATGATGATGATGATGATGATGATGATGATGATGAT 199

```
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J
```

Search completed: June 22, 2003, 22:38:01  
Job time : 228 secs

Search completed: June 22, 2003, 22:38:01  
Job time : 228 secs



XX Fifty nucleic acid molecules encoding human secreted proteins, useful  
PT in the prevention, treatment and diagnosis of cancer, immune disorders,  
PT cardiovascular disorders and neurological diseases -  
XX  
PS Disclosure; Page 415-416; 425pp; English.

XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the  
CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to  
CC AAB34852 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC cytoskeletal; immunosuppressive; neurotropic; neuroprotective; antiviral;  
CC antiallergic; hepatotropic; antidiabetic; antineoplastic; antitumor;  
CC antiviral; anticonvulsant; antibacterial; antifungal; antiparasitic; and  
CC cardiant. The polynucleotides and polypeptides are useful for  
CC preventing, treating or ameliorating a medical condition in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes for  
CC diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the gene  
CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,  
CC agonists and antagonists from the present invention are useful in the  
CC diagnosis, treatment and prevention of cancer, immune disorders,  
CC cardiovascular disorders, wound healing, neurological diseases and  
CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence  
CC used in the exemplification of the present invention.

XX Sequence 205 AA;

Query Match 97.1%; Score 860; DB 21; Length 205;  
Best Local Similarity 84.4%; Pred. No. 7.6e-79;  
Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

OY 1 MSSNNPEYDYLKLLIGDSGVGKSCLLRRADPTYESYSTIGVDFKRTIELDGTI 60  
DB 1 MSSNNPEYDYLKLLIGDSGVGKSCLLRRADPTYESYSTIGVDFKRTIELDGTI 60  
OY 61 KLOI-----ESFNWVKOMLQEIIDRYASENNVKL 88  
DB 61 KLOIWDAGGERFRITISSYRGAGIIVYDVTDOESFNWVKOMLQEIIDRYASENNVKL 120  
OY 89 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGA 148  
DB 121 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGA 180  
OY 149 TAGGAERKNVRIQSTPVKQSGGCC 173  
DB 181 TAGGAERKNVRIQSTPVKQSGGCC 205

RESULT 2  
AAB34844  
ID AAB34844 standard; Protein: 205 AA.

XX AAB34844;

XX 26-JAN-2001 (first entry)

XX Human secreted protein sequence encoded by gene 44 SEQ ID NO:132.

XX Human; secreted protein; diagnosis; cytoskeletal; immunosuppressive;  
KW neurotropic; neuroprotective; antiviral; antineoplastic; hepatotropic;  
KW antidiabetic; antiinflammatory; antitumor; antiviral; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;  
KW cancer; immune disorder; cardiovascular disorder; wound healing;  
KW neurological disease; infectious disease; chromosome identification.  
XX Homo sapiens.

PN WO200058356-A1.

PD 05-OCT-2000.

XX 22-MAR-2000; 2000WO-US07535.

XX 26-MAR-1999; 99US-0126511.

XX 17-DEC-1999; 99US-0172413.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruden SM, Komatsoulis G;

XX WPI: 2000-594639/56.

XX Fifty nucleic acid molecules encoding human secreted proteins, useful  
PT in the prevention, treatment and diagnosis of cancer, immune disorders,  
PT cardiovascular disorders and neurological diseases -

XX Disclosure; Page 416-417; 425pp; English.

XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the  
CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to  
CC AAB34852 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC cytoskeletal; immunosuppressive; neurotropic; neuroprotective; antiviral;  
CC antiallergic; hepatotropic; antidiabetic; antineoplastic; antitumor;  
CC antiviral; anticonvulsant; antibacterial; antifungal; antiparasitic; and  
CC cardiant. The polynucleotides and polypeptides are useful for  
CC preventing, treating or ameliorating a medical condition in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes for  
CC diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the gene  
CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,  
CC agonists and antagonists from the present invention are useful in the  
CC diagnosis, treatment and prevention of cancer, immune disorders,  
CC cardiovascular disorders, wound healing, neurological diseases and  
CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence  
CC used in the exemplification of the present invention.

XX Sequence 205 AA;

Query Match 97.1%; Score 860; DB 21; Length 205;  
Best Local Similarity 84.4%; Pred. No. 7.6e-79;  
Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

OY 1 MSSNNPEYDYLKLLIGDSGVGKSCLLRRADPTYESYSTIGVDFKRTIELDGTI 60  
DB 1 MSSNNPEYDYLKLLIGDSGVGKSCLLRRADPTYESYSTIGVDFKRTIELDGTI 60  
OY 61 KLOI-----ESFNWVKOMLQEIIDRYASENNVKL 88  
DB 61 KLOIWDAGGERFRITISSYRGAGIIVYDVTDOESFNWVKOMLQEIIDRYASENNVKL 120  
OY 89 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGA 148  
DB 121 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGA 180  
OY 149 TAGGAERKNVRIQSTPVKQSGGCC 173  
DB 181 TAGGAERKNVRIQSTPVKQSGGCC 205

RESULT 3  
AAB34816  
ID AAB34816 standard; Protein: 198 AA.

XX AAB34816;

XX 26-JAN-2001 (first entry)  
 XX Human secreted protein sequence encoded by gene 44 SEQ ID NO:104.  
 XX  
 XX Human: secreted protein; diagnosis; cytostatic; immunosuppressive;  
 XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
 XX antidiabetic; antiinflammatory; antitumor; anticonvulsant;  
 XX antidiabetic; antifungal; antiparasitic; cardiant; gene therapy;  
 XX cancer; immune disorder; cardiovascular disorder; wound healing;  
 XX neurological disease; infectious disease; chromosome identification.  
 XX  
 XX Homo sapiens.  
 XX  
 XX W0200058356-A1.  
 XX  
 XX 05-OCT-2000.  
 XX  
 XX 22-MAR-2000; 2000WO-US07535.  
 XX  
 XX 26-MAR-1999; 99US-0126511.  
 XX 17-DEC-1999; 99US-0172413.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Ruben SM, Komatsoulis G;  
 XX WPI: 2000-594639/56.  
 XX N-PSDB; AAC60009.  
 XX  
 XX Fifty nucleic acid molecules encoding human secreted proteins, useful  
 XX in the prevention, treatment and diagnosis of cancer, immune disorders,  
 XX cardiovascular disorders and neurological diseases -  
 XX  
 XX Claim 1; Page 393-394; 425pp; English.  
 XX  
 XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the  
 XX human secreted proteins given in AAB34773 to AAB34823 to  
 XX AAB34852 represent human secreted polypeptide sequences and proteins  
 XX homologous to them, which are given in the exemplification of the present  
 XX invention. Human secreted proteins have activities based on the tissues  
 XX and cells the genes are expressed in. Examples of activities include:  
 XX cytotatic; immunosuppressive; neurotropic; neuroprotective; antiviral;  
 XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 XX antidiabetic; anticonvulsant; antiparasitic; antifungal; antiparasitic; and  
 XX cardiant. The polynucleotides and polypeptides are useful for  
 XX preventing, treating or ameliorating a medical condition in e.g. humans,  
 XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The  
 XX polypeptides can also be used as a food additive or preservative to  
 XX increase or decrease storage capabilities. The polynucleotides are  
 XX useful for chromosome identification. They are also useful as probes for  
 XX diagnosing a disorder related to the female reproductive system,  
 XX particularly breast and/or ovarian cancer. They are also useful in the gene  
 XX therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,  
 XX agonists and antagonists from the present invention are useful in the  
 XX diagnosis, treatment and prevention of cancer, immune disorders,  
 XX cardiovascular disorders, wound healing, neurological diseases and  
 XX infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence  
 XX used in the exemplification of the present invention.  
 XX  
 XX Sequence 198 AA: -  
 XX  
 XX Query Match 92.2%; Score 817; DB 21; Length 198;  
 XX Best Local Similarity 83.8%; Pred. No. 1.6e-74;  
 XX Matches 165; Conservative 0; Mismatches 0; Indels 32; Gaps 1;  
 XX  
 XX 9 DTLFLLIDSDGVGKSCLLRFADDTYTESYSTIGVDFKIRITIELDKTIKIDT----- 64  
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 XX DB 1 DTLFLLIDSDGVGKSCLLRFADDTYTESYSTIGVDFKIRITIELDKTIKIDTMTA 60  
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 XX QY 65 -----ESFNWQWIOLEIDRASNNKLLVNNKDL 96  
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 XX DB 61 GGERRTTSSYRGAHGIIYVDYDQESFNWQWIOLEIDRASNNKLLVNNKDL 120  
 XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 97 TTRKVDYTTAKEPADSLGIPLETSAKNNATNVEOSFMTAAEINKRMGFGATAGAAERS 156  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 121 TTRKVDYTTAKEPADSLGIPLETSAKNNATNVEOSFMTAAEINKRMGFGATAGAAERS 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 157 NVKIOSTPYKOSGGGCC 173  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 181 NVKIOSTPYKOSGGGCC 197  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 RESULT 4  
 ID AAO13525 standard; Protein: 221 AA.  
 XX AAO13525;  
 XX  
 XX AAO13525;  
 XX  
 XX 06-NOV-2001 (first entry)  
 XX  
 XX Human polypeptide SEQ ID NO 27417.  
 XX  
 XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
 XX nervous system disorders; arthritis; inflammation.  
 XX  
 XX Homo sapiens.  
 XX  
 XX W0200164835-A2.  
 XX  
 XX 07-SEP-2001.  
 XX  
 XX 26-FEB-2001; 2001WO-US04927.  
 XX  
 XX 28-FEB-2000; 2000US-0515126.  
 XX 18-MAR-2000; 2000US-0577409;  
 XX (HSE-) HXSEQ INC.  
 XX  
 XX Tang YT, Liu C, Dmanac RT;  
 XX WPI: 2001-514838/56.  
 XX N-PSDB; AA193456.  
 XX  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 XX diagnosing and treating e.g. leukaemia, inflammation and immune  
 XX disorders -  
 XX  
 XX Claim 20; SEQ ID NO 27417; 1399pp + Sequence Listing; English.  
 XX  
 XX The invention relates to human polynucleotides (AA179941-AA193841) and  
 XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
 XX cytokine, cell proliferation or cell differentiation or which may induce  
 XX production of other cytokines in other cell populations. The  
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 XX peptide therapy. The polypeptides have various cytokine-like activities,  
 XX e.g. stem cell growth factor activity, haematopoiesis regulating  
 XX activity, tissue growth factor activity, immunomodulatory activity and  
 XX activity/inhbn activity and may be useful in the diagnosis and/or  
 XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 XX inflammation.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 221 AA:  
 XX  
 XX Query Match 86.7%; Score 768.5; DB 22; Length 221;  
 XX Best Local Similarity 76.5%; Pred. No. 1.5e-69;  
 XX Matches 156; Conservative 8; Mismatches 7; Indels 33; Gaps 2;  
 XX  
 XX 2 SSNNPYDYKIRILLIDSDGVGKSCLLRFADDTYTESYSTIGVDFKIRITIELDKTIK 61  
 XX :||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 XX DB 19 AANAPKIDILKILLIDSDGVGKSCLLRFADDTYTESYSTIGVDFKIRITIELDKTIK 78  
 XX :||||||||||||||||||||||||||||||||||||||||||||||||||||||





CC syndromes such as aplastic anaemia, ischaemic injuries such as myocardial  
 CC infection, stroke, and reperfusion injury, toxin-induced diseases such  
 CC as alcohol-induced liver damage, cirrhosis, and jaundice, wasting  
 CC diseases such as cachexia, viral infections, and osteoporosis. They can  
 CC also be used to stimulate cell proliferation for use in transplantation  
 CC or to produce cells to fight an infection or a cancer or to correct a  
 CC genetic defect in a disease such as sickle cell beta thalassaemia, cystic  
 CC fibrosis or Huntington's chorea. Antagonists can be used to prevent or  
 CC treat a disorder associated with cell proliferation e.g. cancers or  
 CC inflammation, e.g. Addison's disease, adult respiratory distress  
 CC syndrome, allergies, asthma, atherosclerosis, bronchitis, cholecystitis,  
 CC Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis,  
 CC diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis,  
 CC gout, Graves' disease, hypercalcaemia, irritable bowel syndrome, lupus  
 CC erythematosus, multiple sclerosis, myasthenia gravis, inflammation,  
 CC osteoarthritis, osteoporosis, pancreatitis, polyarthritis, rheumatoid  
 CC arthritis, scleroderma, Sjogren's syndrome, and autoimmune thyroiditis,  
 CC complications of cancer, haemodialysis, extracorporeal circulation,  
 CC infections and trauma.

XX Sequence 201 AA:  
 SQ

Query Match 86.5%; Score 766.5; DB 20; Length 201;  
 Best Local Similarity 77.2%; Pred. No. 2,1e-69;  
 Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

OY 4 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYSTIGVDFKRIEIDGKTIKQ 63  
 DB 1 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYSTIGVDFKRIEIDGKTIKQ 60  
 OY 64 I-----ESFNWVKOMLOEIDRYASNNKLLVG 91  
 DB 61 IMDTAGGERFRITTSYRGAGHGIIVYDVTQDESANVWOMLOEIDRYASNNKLLVG 120  
 OY 92 NKCDLTKRKVYDNTTAKRFADSGIPFLETSKNNATNVEOSFPMMAEIKRMGPGATAG 151  
 DB 121 NKSDLTTRKRVYDNTTAKRFADSGIPFLETSKNNATNVEOAFPTMAEIKRMGPGATAG 180  
 OY 152 GAESNNVTKIOSTPYKSGGGCC 173  
 DB 181 G-ERPNKAKIDSTPYKPAAGGCC 201

RESULT 7  
 AAU28024  
 ID AAU28024 standard; Protein: 201 AA.  
 AC AAU28024;  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secretory protein, Seq ID No 193.  
 XX  
 KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
 KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
 KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
 KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
 KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
 KW gut protection; lung; liver fibrosis; immune deficiency; infection;  
 KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
 KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
 KW fertility; analgesic; pain; antigen.

XX Homo sapiens.  
 OS  
 XX MO20016689-A2.  
 PN  
 XX 13-SEP-2001.  
 PD  
 XX 05-MAR-2001; 2001MO-US04942.  
 PF  
 XX 07-MAR-2000; 2000US-0519705.  
 PR  
 XX 19-MAY-2000; 2000US-0574454.  
 PR

PR 17-JUN-2000; 2000US-0536193.  
 PR 14-JUL-2000; 2000US-0616847.  
 PR 19-SEP-2000; 2000US-0665363.  
 PR 20-OCT-2000; 2000US-0693267.  
 XX  
 PA (HXSE-) HXSBQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;  
 PI Zhao QA, Yang Y, Dmanac RT, Zhang J, Chen R, Xue AJ, Wang J;  
 XX  
 DR WPI: 2001-589934/66.  
 DR N-PSDB: AAS44924.  
 XX  
 PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
 PT prepared from various human tissues, for diagnosis and treatment of  
 PT cancer, neurological, inflammatory, and autoimmune disorders -  
 XX  
 PS Example 3; SEQ ID No 193; 107pp; English.

XX The invention relates to novel isolated human secreted polypeptides (I)  
 CC and polynucleotides (II). (I) and (II) are useful for treating  
 CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,  
 CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is  
 CC involved in increasing haematopoiesis, stem cell survival, bone growth  
 CC and remodeling. (II). (II) and modulators of (II) are useful for  
 CC prophylaxis or treatment of one or more cancers. (II) is also useful for  
 CC creating transgenic animals useful for studying the in vivo activities of  
 CC the polypeptide as well as for studying modulators of the polypeptides.  
 CC (I) induces the proliferation of neural cells and regeneration of nerve  
 CC and brain tissue and is useful for the treatment of central and  
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
 CC sclerosis. In addition (I) is involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis and is useful for treating myeloid  
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
 CC tissue growth and in tissue repair, healing of burns, incisions,  
 CC ulcers or treating osteoporosis, osteoarthritis, bone degenerative  
 CC disorders or periodontal disease. Furthermore, (I) is also useful for  
 CC gut protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, various immune deficiencies and  
 CC disorders including severe combined immunodeficiency (SCID), bacterial or  
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
 CC reactions and conditions, such as asthma or other respiratory problems.  
 CC In addition, (I) affects biorhythms or circadian cycles of rhythms,  
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
 CC analgesic effects or other pain reducing effects, immunoglobulin like  
 CC activity and can act as an antigen in a vaccine composition to raise an  
 CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
 CC amino acid sequences of the invention.

XX Sequence 201 AA:  
 SQ

Query Match 86.5%; Score 766.5; DB 22; Length 201;  
 Best Local Similarity 77.2%; Pred. No. 2,1e-69;  
 Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

OY 4 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYSTIGVDFKRIEIDGKTIKQ 63  
 DB 1 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYSTIGVDFKRIEIDGKTIKQ 60  
 OY 64 I-----ESFNWVKOMLOEIDRYASNNKLLVG 91  
 DB 61 IMDTAGGERFRITTSYRGAGHGIIVYDVTQDESANVWOMLOEIDRYASNNKLLVG 120  
 OY 92 NKCDLTKRKVYDNTTAKRFADSGIPFLETSKNNATNVEOSFPMMAEIKRMGPGATAG 151  
 DB 121 NKSDLTTRKRVYDNTTAKRFADSGIPFLETSKNNATNVEOAFPTMAEIKRMGPGATAG 180  
 OY 152 GAESNNVTKIOSTPYKSGGGCC 173  
 DB 181 G-ERPNKAKIDSTPYKPAAGGCC 201

Db	181 G-ERPNTKIDSTPVPKAGGCC 201
Db	181 SATD-MSKVKIDGSRPVETKNSGCC 205
XX	ABBS59808
XX	ABBS59808 standard; Protein; 205 AA.
XX	ABBS59808;
XX	26-MAR-2002 (first entry)
XX	Drosophila melanogaster polypeptide SEQ ID NO 6216.
XX	Drosophila: developmental biology; cell signalling; insecticide;
XX	pharmaceutical.
XX	Drosophila melanogaster.
XX	MO200171042-A2.
XX	27-SEP-2001.
XX	23-MAR-2001; 2001WO-US09231.
XX	23-MAR-2000; 2000US-191637P.
XX	11-JUL-2000; 2000US-0614150.
XX	(PEKE ) PE CORP NY.
XX	Venter JC, Adams M, Li PWD, Myers EW;
XX	WPI; 2001-6556860/75.
XX	N-PSDB; ABL03911.
XX	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX	Disclosure; SEQ ID NO 6216; 21pp + Sequence Listing; English.
XX	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB101840-AB130511), expressed DNA sequences (AB857737-AB872072).
XX	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX	Sequence 205 AA;
XX	Query Match 77.1%; Score 683; DB 22; Length 205; Best Local Similarity 69.4%; Pred. No. 6e-61; Mismatches 17; Indels 34; Gaps 3
XX	Matches 143; Conservative 12;
QY	1 MSNNPEEDYLFKLLIGDSGVKSCLLRFADDDYTSYSTIGVDFKRIETLDGKTI 60
Db	1 MSSVNPEDYLFKLLIGDSGVKSCLLRFADDDYTSYSTIGVDFKRIETLDGKTI 60
QY	61 KIQI-----ESFNMYKWLQETIRYASENVNLT 88
Db	61 KIQIWDTAGGERFRITTSYRGAHGIIVYDCDTQDESFNNYKWLLEETIERACENVNLT 120
QY	89 LVGKCDLTTKKVDDYTAKEPADSLGIPLETSAKNAITNVQSPWTMAAETIKKMGPCA 148
Db	121 LVGKNSDITTTKKVDDHTAAEAQOLGIPLETSAKSATNVEQAFMTMAAEIKKNVGPSS 180
QY	149 TAGAAEKSNVKI-OSTPVKOSGGGCC 173
Db	181 SATD-MSKVKIDGSRPVETKNSGCC 205

RESULT	9
AAAG07763	
ID	AAG07763 standard; Protein; 202 AA.
XX	
AC	AAG07763;
XX	
DT	17-OCT-2000 (first entry)
DE	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 5040
XX	
KW	Protein identification; signal transduction pathway; n
KM	hybridisation assay; genetic mapping; gene expression
RW	termination sequence.
XX	
OS	Arabidopsis thaliana.
PN	EPI033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130892.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	14-MAY-1999; 99US-0134766.
PR	18-MAY-1999; 99US-0134941.
PR	19-MAY-1999; 99US-0135124.
PR	20-MAY-1999; 99US-0135353.
PR	21-MAY-1999; 99US-0135629.
PR	24-MAY-1999; 99US-0136021.
PR	25-MAY-1999; 99US-0136392.
PR	27-MAY-1999; 99US-0136782.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139819.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 20-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 29-SEP-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157753.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158232.  
PR 13-OCT-1999; 99US-0158233.  
PR 13-OCT-1999; 99US-0158234.  
PR 13-OCT-1999; 99US-0158235.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 71 9%; Score 637; DB 21; Length 202;  
Beat Local Similarity 65.3%; Pred No. 2, 6e-56;  
Matches 133; Conservative 11; Mismatches 25; Indels 34; Gaps 3;

QY 4 MNEPDYLFKLLIGSGGKSCILLRPAADYTBESYISTGVDFKRTIELDGTIKLQ 63  
DB 1 MNEPDYLFKLLIGSGGKSCILLRPAADSTLSTLSTGVDFKRTIELDGTIKLQ 60  
QY 64 I-----ESTNNVKOMLOEIDIRVASENVKLLVG 91  
DB 61 IWDYGAQGEFRITTSYRKAGIIVYVDYDLESTNNVKOMLEIDIRVASENVKLLVG 120  
QY 92 NKCDLITKKVVDYTTAKEPADSLGIPLETSAANAATNVQSFMTMAAEIKRKWGCGATAG 151



PR 13-AUG-1999: 99US-0148565.  
 PR 16-AUG-1999: 99US-0148684.  
 PR 17-AUG-1999: 99US-0149368.  
 PR 18-AUG-1999: 99US-0149375.  
 PR 19-AUG-1999: 99US-0149426.  
 PR 20-AUG-1999: 99US-0149722.  
 PR 20-AUG-1999: 99US-0149723.  
 PR 20-AUG-1999: 99US-0149929.  
 PR 23-AUG-1999: 99US-0149902.  
 PR 23-AUG-1999: 99US-0149930.  
 PR 25-AUG-1999: 99US-0150566.  
 PR 26-AUG-1999: 99US-0150884.  
 PR 27-AUG-1999: 99US-0151065.  
 PR 27-AUG-1999: 99US-0151066.  
 PR 30-AUG-1999: 99US-0151080.  
 PR 31-AUG-1999: 99US-0151303.  
 PR 01-SEP-1999: 99US-0151930.  
 PR 07-SEP-1999: 99US-0152363.  
 PR 10-SEP-1999: 99US-0153070.  
 PR 13-SEP-1999: 99US-0153758.  
 PR 15-SEP-1999: 99US-0154018.  
 PR 16-SEP-1999: 99US-0154039.  
 PR 20-SEP-1999: 99US-0154779.  
 PR 22-SEP-1999: 99US-0155139.  
 PR 23-SEP-1999: 99US-0155486.  
 PR 24-SEP-1999: 99US-0155659.  
 PR 28-SEP-1999: 99US-0156458.  
 PR 29-SEP-1999: 99US-0156596.  
 PR 04-OCT-1999: 99US-0157117.  
 PR 05-OCT-1999: 99US-0157753.  
 PR 06-OCT-1999: 99US-0157865.  
 PR 07-OCT-1999: 99US-0158029.  
 PR 08-OCT-1999: 99US-0158232.  
 PR 12-OCT-1999: 99US-0158369.  
 PR 13-OCT-1999: 99US-0159283.  
 PR 13-OCT-1999: 99US-0159293.  
 PR 13-OCT-1999: 99US-0159329.  
 PR 14-OCT-1999: 99US-0159330.  
 PR 14-OCT-1999: 99US-0159331.  
 PR 14-OCT-1999: 99US-0159337.  
 PR 14-OCT-1999: 99US-0159637.  
 PR 14-OCT-1999: 99US-0159638.  
 PR 18-OCT-1999: 99US-0159584.  
 PR 21-OCT-1999: 99US-0160741.  
 PR 21-OCT-1999: 99US-0160767.  
 PR 21-OCT-1999: 99US-0160768.  
 PR 21-OCT-1999: 99US-0160770.  
 PR 21-OCT-1999: 99US-0160814.  
 PR 21-OCT-1999: 99US-0160815.  
 PR 22-OCT-1999: 99US-0160980.  
 PR 22-OCT-1999: 99US-0160981.  
 PR 22-OCT-1999: 99US-0160989.  
 PR 25-OCT-1999: 99US-0161404.  
 PR 25-OCT-1999: 99US-0161405.  
 PR 25-OCT-1999: 99US-0161406.  
 PR 26-OCT-1999: 99US-0161359.  
 PR 26-OCT-1999: 99US-0161360.  
 PR 26-OCT-1999: 99US-0161361.  
 PR 28-OCT-1999: 99US-0161920.  
 PR 28-OCT-1999: 99US-0161992.  
 PR 28-OCT-1999: 99US-0161993.  
 PR 29-OCT-1999: 99US-0162142.

Query Match 71.4% Score 633: DB 21: Length 202;  
 Best Local Similarity 65.8% Pred. No 6, 7e-56;  
 Matches 133; Conservative 10; Mismatches 25; Indels 34; Gaps 3;  
 OY 4 NMPEDYFKLLIGSGVSKSLLPADDTYTESYISTGVDFKRTIEDGKTIKQ 63  
 DB 1 NMPEDYFKLLIGSGVSKSLLPADDTYTESYISTGVDFKRTIEDGKTIKQ 60  
 OY 64 I-----ESFNNAKQMLQELDRASENKKLLVG 91

DB 61 IWDPAQGEFRTITSSYRCAGIIVTIDVDESFNNNAQWMLNEIDRTASENKKLLVG 120  
 OY 92 NKCDLTKKVVDDYTTAKEFADSLGIPLETSAKKNAIIVDEOSFMTAAEIKRKGPGATAG 151  
 DB 121 NKNDLTSOKVSTFETAAAFADSLGIPLETSAKKNAIIVDEAMATTAIKTRAA-SOPAG 179  
 OY 152 GAESKNVKIQTSTPVKQSGGGCC 173  
 DB 180 GAKPFTVQIRGQPVNQO-SGCC 200  
 RESULT 11  
 AAG30498  
 ID AAG30498 standard; Protein: 258 AA.  
 AC AAG30498;  
 XX  
 DT 17-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment seq ID NO: 36472.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 PD 06-SEP-2000.  
 XX  
 PR 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0122180.  
 PR 09-MAR-1999; 99US-0122548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134961.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136932.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139493.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144886.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148365.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161369.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 69.9%; Score 619.5; DB 21; Length 258;  
 Best Local Similarity 61.6%; Pred. No. 2,le-54;  
 Matches 125; Conservative 15; Mismatches 30; Indels 33; Gaps 2;

Oy	3	SANEDYLFKLLILGDSGVKSCLLLRADPTYESISTGVDPKIRTEILGKTIK	62	PR	28-MAY-1999	99US-0136788
				PR	18-JUN-1999	99US-0137222
Db	55	TNNEVDYLFKLLILGDSGVKSCLLLRSDSYESTISTGVDPKIRTEILGKTIK	114	PR	03-JUN-1999	99US-0137528
				PR	18-JUN-1999	99US-0137502
Oy	63	QI-----ESFRNKKOMLQEDIRYASEVNNKLLY	90	PR	07-JUN-1999	99US-0137724
				PR	04-JUN-1999	99US-0137502
Db	115	QIMPTAQGEFRKTISSYKAGAGIIVYDVTDESEFNNKKOMLSTIDRKASDVNNKLLY	174	PR	08-JUN-1999	99US-0138094
				PR	10-JUN-1999	99US-0138540
Oy	91	GNCKDILTKKVVDDYTTANEPADSLGIPLETISAKNATINVEDQSPYMAAEIRKRMGPQATA	150	PR	14-JUN-1999	99US-0138847
				PR	16-JUN-1999	99US-0139452
Db	175	GNKSDILLENALPIETETAFNAFDEIGLIPFMEITSADATINVEDQAFMAASISIKERMASOPAG	234	PR	16-JUN-1999	99US-0139453
				PR	17-JUN-1999	99US-0139454
Oy	151	GAEKSNVKIQSTPVKQSGGCC	173	PR	18-JUN-1999	99US-0139454
				PR	18-JUN-1999	99US-0139455
Db	235	NNARPPTVQIKQAPVAK-NCCC	256	PR	18-JUN-1999	99US-0139457
				PR	18-JUN-1999	99US-0139458
				PR	18-JUN-1999	99US-0139459
RESULT 12				PR	18-JUN-1999	99US-0139460
AAG30499				PR	18-JUN-1999	99US-0139461
XX	AAG30499	standard; Protein; 203 AA.		PR	18-JUN-1999	99US-0139462
XX	AAG30499			PR	18-JUN-1999	99US-0139463
AC				PR	18-JUN-1999	99US-0139750
XX	17-OCT-2000	(first entry)		PR	21-JUN-1999	99US-0139763
XX				PR	21-JUN-1999	99US-0139817
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 36473.		PR	22-JUN-1999	99US-0139899
XX				PR	23-JUN-1999	99US-0140353
XX				PR	23-JUN-1999	99US-0140354
AC				PR	24-JUN-1999	99US-0140693
XX	17-OCT-2000	(first entry)		PR	28-JUN-1999	99US-0140823
XX				PR	29-JUN-1999	99US-0140991
FM	Protein identification: signal transduction pathway; metabolic pathway;			PR	30-JUN-1999	99US-0141287
KM	hybridisation assay; genetic mapping; gene expression control; promoter;			PR	01-JUL-1999	99US-0141842
XX	termination sequence.			PR	01-JUL-1999	99US-0142154
XX	Arabidopsis thaliana.			PR	02-JUL-1999	99US-0142055
OS	EP1033405-A2.			PR	06-JUL-1999	99US-0142390
PN	06-SEP-2000.			PR	08-JUL-1999	99US-0142803
XX				PR	09-JUL-1999	99US-0142920
XX	25-FEB-2000; 2000EP-0301439.			PR	12-JUL-1999	99US-0142977
XX				PR	13-JUL-1999	99US-0143542
XX	25-FEB-1999; 99US-0121823.			PR	14-JUL-1999	99US-0143624
PR	05-MAR-1999; 99US-0123180.			PR	15-JUL-1999	99US-0144005
PR	09-MAR-1999; 99US-0123548.			PR	16-JUL-1999	99US-0144085
PR	23-MAR-1999; 99US-0125788.			PR	19-JUL-1999	99US-0144325
PR	25-MAR-1999; 99US-0126264.			PR	19-JUL-1999	99US-0144331
PR	29-MAR-1999; 99US-0126785.			PR	19-JUL-1999	99US-0144332
PR	01-APR-1999; 99US-0127462.			PR	19-JUL-1999	99US-0144333
PR	06-APR-1999; 99US-0128234.			PR	19-JUL-1999	99US-0144334
PR	08-APR-1999; 99US-0128714.			PR	19-JUL-1999	99US-0144335
PR	16-APR-1999; 99US-0129845.			PR	19-JUL-1999	99US-0144336
PR	19-APR-1999; 99US-0130077.			PR	20-JUL-1999	99US-0144337
PR	21-APR-1999; 99US-0130449.			PR	20-JUL-1999	99US-0144338
PR	23-APR-1999; 99US-0130510.			PR	20-JUL-1999	99US-0144339
PR	28-APR-1999; 99US-0131449.			PR	20-JUL-1999	99US-0144340
PR	30-APR-1999; 99US-0132048.			PR	21-JUL-1999	99US-0144341
PR	30-APR-1999; 99US-0132407.			PR	21-JUL-1999</	

PR	04-ANG-1999	9905-0147302
PR	05-ANG-1999	9905-0147260
PR	05-AUG-1999	9905-0147762
PR	06-ANG-1999	9905-0147703
PR	06-AUG-1999	9905-0147416
PR	09-ANG-1999	9905-0147593
PR	09-ANG-1999	9905-0147935
PR	10-ANG-1999	9905-0148171
PR	11-ANG-1999	9905-0148341
PR	12-ANG-1999	9905-0148319
PR	13-ANG-1999	9905-0148565
PR	16-ANG-1999	9905-0149368
PR	17-ANG-1999	9905-0149175
PR	18-ANG-1999	9905-0149426
PR	20-ANG-1999	9905-0149722
PR	20-ANG-1999	9905-0149723
PR	20-ANG-1999	9905-0149902
PR	23-ANG-1999	9905-0149929
PR	23-ANG-1999	9905-0149930
PR	23-ANG-1999	9905-0150566
PR	26-ANG-1999	9905-0150884
PR	27-ANG-1999	9905-0151065
PR	27-ANG-1999	9905-0151066
PR	27-ANG-1999	9905-0151080
PR	30-ANG-1999	9905-0151303
PR	31-ANG-1999	9905-0151338
PR	01-SEP-1999	9905-0151363
PR	07-SEP-1999	9905-0152370
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154039
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779
PR	22-SEP-1999	9905-0155133
PR	22-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0156559
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157573
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158362
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	13-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159531
PR	14-OCT-1999	9905-0159637
PR	18-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159684
PR	21-OCT-1999	9905-0160584
PR	21-OCT-1999	9905-0160741
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160982
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161350
PR	26-OCT-1999	9905-0161360
PR	28-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161992
PR	28-OCT-1999	9905-0161993
PR	29-OCT-1999	9905-0162142

	Query Match	69.8%;	Score 618.5;	DB 21:	Length 203;	
	Best Local Similarity	61.9%;	Pred. No. 2e-54;	Mismatches 125;	Conservative 14;	Indels 33; Gaps 2
OY	4 MNPEVDYFFKLLIDSDGVGSKCLLRFADPDVTFYESYSTIGVDFFKRTIELDGKTIRQ	63				
Db	1 MNPEVDYFFKLLIDSDGVGSKCLLRFSDSDYESISTIGVDFFKRIVEODGKTIRQ	60				
OY	64 I-----ESFNWVKOMIQEIDRYASENVNKLWG	91				
Db	61 IWDTAGGERFTITSSYYKGAGHIIIVDYDTEDEFNNYKOWMLSEIDRYASNVNKLWG	120				
OY	92 NKCDDTFRKKVVDTYTAKEFADSLGIPELETSAKNATNVQSPMTAAEIKTKRMGPQATAG	151				
Db	121 NKSDLTENNAIIFYETAKAFPADEIGLPFMETSAKDATNVQAPMAASAIKERNASOPAGN	180				
OY	152 GAEKSNVRIQSTPVKOSGGGCC	173				
Db	181 NARPTVOIRGQPVAOK-NGCC	201				
RESULT 13						
ID	AAG30710					
XX	AAG30710 standard; Protein; 218 AA.					
AC	AAG30710;					
XX						
DT	17-OCT-2000 (first entry)					
XX						
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 36764.					
XX						
KM	Protein identification; signal transduction pathway; metabolic pathway;					
KW	hybridisation assay; genetic mapping; gene expression control; promoter;					
KX	termination sequence.					
XX						
OS	Arabidopsis thaliana.					
XX						
PN	EPI033405-A2.					
XX						
PD	06-SEP-2000.					
XX						
PF	25-FEB-2000; 2000EP-0301439.					
XX						
PR	25-FEB-1999; 99US-0121825.					
PR	05-MAR-1999; 99US-0123180.					
PR	09-MAR-1999; 99US-0122548.					
PR	23-MAR-1999; 99US-0125788.					
PR	25-MAR-1999; 99US-0126264.					
PR	29-MAR-1999; 99US-0126785.					
PR	01-APR-1999; 99US-0127462.					
PR	06-APR-1999; 99US-0128234.					
PR	08-APR-1999; 99US-0128714.					
PR	16-APR-1999; 99US-0128845.					
PR	19-APR-1999; 99US-0130077.					
PR	21-APR-1999; 99US-0130449.					
PR	23-APR-1999; 99US-0130510.					
PR	23-APR-1999; 99US-0130891.					
PR	28-APR-1999; 99US-0131449.					
PR	30-APR-1999; 99US-0132048.					
PR	30-APR-1999; 99US-0132407.					
PR	04-MAY-1999; 99US-0132484.					
PR	05-MAY-1999; 99US-0132485.					
PR	06-MAY-1999; 99US-0132486.					
PR	06-MAY-1999; 99US-0132487.					
PR	07-MAY-1999; 99US-0132863.					
PR	11-MAY-1999; 99US-0134256.					
PR	14-MAY-1999; 99US-0134218.					
PR	14-MAY-1999; 99US-0134219.					
PR	14-MAY-1999; 99US-0134221.					
PR	14-MAY-1999; 99US-0134370.					
PR	18-MAY-1999; 99US-0134768.					
PR	19-MAY-1999; 99US-0134941.					



PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146386.
PR 21-MAY-1999;	99US-0135253.	PR 02-AUG-1999;	99US-0146388.
PR 24-MAY-1999;	99US-0135629.	PR 02-AUG-1999;	99US-0146389.
PR 25-MAY-1999;	99US-0136021.	PR 03-AUG-1999;	99US-0147038.
PR 27-MAY-1999;	99US-0136392.	PR 04-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147319.
PR 03-JUN-1999;	99US-0137502.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 10-AUG-1999;	99US-0147935.
PR 14-JUN-1999;	99US-0138847.	PR 11-AUG-1999;	99US-0148171.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 17-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149272.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149299.
PR 18-JUN-1999;	99US-0139459.	PR 23-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139461.	PR 25-AUG-1999;	99US-0150326.
PR 18-JUN-1999;	99US-0139462.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139463.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139817.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140053.	PR 31-AUG-1999;	99US-0151348.
PR 23-JUN-1999;	99US-0140054.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140095.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153707.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 02-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 06-JUL-1999;	99US-0142290.	PR 22-SEP-1999;	99US-0155139.
PR 08-JUL-1999;	99US-0142803.	PR 23-SEP-1999;	99US-0155486.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143624.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 12-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 13-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159337.
PR 21-JUL-1999;	99US-0144884.	PR 18-OCT-1999;	99US-0159837.
PR 21-JUL-1999;	99US-0145086.	PR 18-OCT-1999;	99US-0159838.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160741.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160767.
PR 23-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160768.
PR 23-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160770.
PR 23-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160981.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0160989.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161933.  
 PR 29-OCT-1999; 99US-0162142.  
 Query Match 69.1%; Score 612; DB 21; Length 218;  
 Best Local Similarity 65.0%; Pred. No. 9,8e-54;  
 Matches 128; Conservative 11; Mismatches 24; Indels 34; Gaps 3;  
 OY 9 DYLFKLLIGDSGVKSCLLRFADDTTESTISTIGDFKRTTELGKTKLIDT----- 64  
 DB 22 DYLFKLLIGDSGVKSCLLRFADDTSTIDSTISTIGDFKRTTELGKTKLIDT----- 81  
 OY 65 -----ESFNNVQMLQELIDRYASENNKLLVGNKCDL 96  
 DB 82 GOERFRTTSYRGAAGLIYVDTLDSEFNNVQMLNEIDRYASENNKLLVGNKCDL 141  
 OY 97 TTKKVVDTTAKEPDSTGIPPLETSANRNATNVEDSEFNTMAAEIKKRGPGATGAGAEKS 156  
 DB 142 TSOKVSTETAKAFDEGIPPLETSANRNATNVEEFAMTAIRTRMA-SQFAGSKRP 200  
 OY 157 NKIGSTPVKSGGSC 173  
 DB 201 TVQIRGQPVNOQ-SGCC 216  
 RESULT 14  
 AAG30709  
 ID AAG30709 standard; Protein; 221 AA.  
 AC AAG30709;  
 DT 17-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 36763.  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 OS Arabidopsis thaliana.  
 XX  
 XX EPI033405-A2.  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132404.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140333.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144684.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.

[illegible]

PR 06-MAY-1999;	99US-0132486.
PR 06-MAY-1999;	99US-0132487.
PR 11-MAY-1999;	99US-0132863.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145087.
<hr/>	
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147199.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 10-AUG-1999;	99US-0147935.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150888.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151428.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.

PR 21-OCT-1999; 990S-0160814.  
 PR 21-OCT-1999; 990S-0160815.  
 PR 22-OCT-1999; 990S-0160980.  
 PR 22-OCT-1999; 990S-0160981.  
 PR 22-OCT-1999; 990S-0160989.  
 PR 25-OCT-1999; 990S-0161404.  
 PR 25-OCT-1999; 990S-0161405.  
 PR 25-OCT-1999; 990S-0161406.  
 PR 26-OCT-1999; 990S-0161359.  
 PR 26-OCT-1999; 990S-0161360.  
 PR 26-OCT-1999; 990S-0161361.  
 PR 28-OCT-1999; 990S-0161920.  
 PR 28-OCT-1999; 990S-0161992.  
 PR 28-OCT-1999; 990S-0161993.  
 PR 29-OCT-1999; 990S-0162142.

Query Match 69.18; Score 612; DB 21; Length 254;  
 Best Local Similarity 65.08; Pred. No. 1,2e-53;  
 Matches 128; Conservative 11; Mismatches 24; Indels 34; Gaps 3;

QY 9 DYFRLILGDSGVGKSCILIRPADDTESYTGVDKFKRTIELDQTKIKOI---- 64  
 Db 58 DYFRLILGDSGVGKSCILIRPADDTESYTGVDKFKRTIELDQTKIKOIIMDTA 117  
 QY 65 -----ESPNNYKOHLOIDIRYASENVNKLIVGNKCDL 96  
 Db 118 GDERFRITTSYRGAGHGIIVTYVDLESFPNNYKOHLEIDRYASENVNKLIVGNKCDL 177  
 QY 97 TTKKVVDTYTAKEPADSLGIPLETSAKNATUNVQSFMTMAELIKRKGPGATAGGAERK 156  
 Db 178 TSKRVSTETAKAFADSLGIPLETSAKNATUNVQSFMTMAELIKRKGPGATAGGAERK 236  
 QY 157 NVKIGSTPVKQSGGCC 173  
 Db 237 TVQIRGQPVNOQ-SGCC 232

Search completed: June 18, 2003, 15:21:09  
 Job time : 72 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using SW model

Run on: June 18, 2003, 15:19:09 ; Search time 41 Seconds  
(without alignments)  
405,640 Million cell updates/sec

Title: US-09-820-003b-2

Perfect score: 886  
Sequence: 1 MSSNPEYDYLKLLIGDS.....EKSNNKIQSPVKGSGGCC 173

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	860	97.1	205	1 TVH0YP	GTP-binding protein
2	860	97.1	205	1 TVD0YP	GTP-binding protein
3	860	97.1	205	1 TVW0YP	GTP-binding protein
4	849	95.8	205	1 TVR0YP	GTP-binding protein
5	812.5	91.7	201	2 D38625	GTP-binding protein
6	771.5	87.1	201	2 S06147	GTP-binding protein
7	733	82.7	205	2 S38339	GTP-binding protein
8	671	75.7	202	2 J80318	GTP-binding protein
9	660	74.5	205	2 T33781	GTP-binding protein
10	658.5	74.3	203	2 S10096	GTP-binding protein
11	646.5	73.0	203	2 J04105	GTP-binding protein
12	646.5	73.0	203	2 T50322	GTP-binding protein
13	642.5	72.5	203	2 J03247	GTP-binding protein
14	631.5	71.5	201	2 UC5327	GTP-binding protein
15	627.5	70.8	203	2 S34253	GTP-binding protein
16	623.5	70.4	206	2 S04599	GTP-binding protein
17	619.5	69.9	226	2 B86133	GTP-binding protein
18	618.5	69.9	202	2 S41430	GTP-binding protein
19	618.5	69.8	202	2 B38202	GTP-binding protein
20	613.5	69.2	202	2 S38740	GTP-binding protein
21	612	69.1	221	2 H71444	GTP-binding protein
22	598	67.5	218	2 T07609	GTP-binding protein
23	590.5	66.6	202	2 S72515	GTP-binding protein
24	587	66.3	201	2 S39565	GTP-binding protein
25	574	64.8	206	2 T14391	GTP-binding protein
26	572.5	64.6	196	2 PS0279	GTP-binding protein
27	571	64.4	206	1 TVB0Q2	GTP-binding protein
28	523	59.0	208	2 A38202	GTP-binding protein
29	468.5	53.9	208	2 A34716	GTP-binding protein

30	461	52.0	203	2 B34716	GTP-binding protein
31	447	50.5	216	2 T48378	GTP-binding protein
32	442	49.9	215	2 T14565	GTP-binding protein
33	439.5	49.6	215	2 S57478	GTP-binding protein
34	438.5	49.5	216	2 J50640	GTP-binding protein
35	430.5	48.2	200	2 S12790	GTP-binding protein
36	429.5	48.5	215	2 S57462	GTP-binding protein
37	429.5	48.5	216	2 S57471	GTP-binding protein
38	429.5	48.5	222	2 T14405	GTP-binding protein
39	428.5	48.4	216	2 T45901	GTP-binding protein
40	426.5	48.1	215	2 S57474	GTP-binding protein
41	425.5	48.0	216	2 S33900	GTP-binding protein
42	418	47.2	207	2 B49647	GTP-binding protein
43	418	47.2	207	2 B36364	GTP-binding protein
44	410.5	46.3	206	2 T78851	GTP-binding protein
45	409.5	46.2	209	2 B38625	GTP-binding protein

## ALIGNMENTS

## RESULT 1

GTP-binding protein Rab1 - human

N:Alternate names: protein DKFZ564B163.1

C:Species: Homo sapiens (man)

C>Date: 29-Jun-1990 #sequence\_rev1sion 06-Dec-1996 #text\_change 19-Jan-2001

C:Accession: A34323; T08698

R:Zahrouk, A.; Touchot, N.; Chardin, P.; Tavillan, A.

J. Biol. Chem. 264, 12394-12401, 1989

A:Title: The human Rab genes encode a family of GTP-binding proteins related to yeast

A:Reference number: A34323; MUID:89308668; PMID:2501306

A:Accession: A34323

A:Molecule type: mRNA

A:Residues: 1-205 <ZMH>

A:Cross-references: GB:J04941; GB:M28209; MID:9550059; PIDN:AAA60240.1; PID:9550060

A:Experimental source: phaeochromocytoma

R:Wambolt, R.; Heubner, D.; Meves, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, March 1999

A:Reference number: 216471

A:Accession: T08698

A:Molecule type: mRNA

A:Residues: 1-64,141-205 <MAN>

A:Cross-references: EMBL:AL050268

A:Experimental source: fetal brain, clone DKFZ564B163

C:Genetics:

A:Gene: GDB:RAB1

A:Cross-references: GDB:118857; OMIM:179508

A:Map position: 4p15.31-4p15.31

A:Note: DKFZ564B163.1

C:Function:

A:Description: probably involved in protein transport from the endoplasmic reticulum

A:Superfamily: ras transforming protein; translation elongation factor Tu homolog

C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipid protein; membran

F:1-205/Product: GTP-binding protein Rab1 #status predicted <MAN>

F:1-64,141-205/Product: GTP-binding protein Rab1, splice variant #status predicted <M>  
F:12-117/Domain: translation elongation factor Tu homolog <ETU>  
F:18-23/Region: nucleotide-binding motif A (P-loop)  
F:62-67/Region: nucleotide-binding motif B  
F:124-127/Region: GTP-binding NKXD motif  
F:154-156/Region: GTP-binding SAKL motif  
F:24,25,43,124,125,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #  
F:194/Binding site: phosphate (Ser) (covalent) #status predicted  
F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status experimental

Query Match 97.1% Score 860; DB 1; Length 205;

Best local similarity 84.4% Pred. No. 2,2e-62;

Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

DB 1 MSSNPEYDYLKLLIGDSGVKSCILLRFADDTYESTYSTIGVDFIRIELDKRTI 60  
|||||  
1 MSSNPEYDYLKLLIGDSGVKSCILLRFADDTYESTYSTIGVDFIRIELDKRTI 60

QY 61 KIQI-----ESFNNYKQWLOEIDRYASENVKL 88  
 |||||  
 Db 61 KQIWDTAGQERFRTITSSYRGAGIIVYDVDTQESFNNYKQWLOEIDRYASENVKL 120  
 QY 89 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKKRMGPGA 148  
 Db 121 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKKRMGPGA 180  
 QY 149 TAGGAEKSNVKIQSTPVRKSGGCC 173  
 |||||  
 Db 181 TAGGAEKSNVKIQSTPVRKSGGCC 205

## RESULT 2

YVDGYP  
 C:Species: Canis lupus familiaris (dog)  
 C:Date: 19-Feb-1994 #sequence\_revision 06-Dec-1996 #text\_change 19-Jan-2001  
 C:Accession: S19104; A36364; S15600  
 R:Zerial, M.  
 submitted to the EMBL Data Library, August 1990  
 A:Reference number: S19104  
 A:Accession: S19104  
 A:Molecule type: mRNA  
 A:Residues: 1-205 <ZER>  
 A:Cross-references: EMBL:X56384  
 R:Chavrier, P.; Vignon, M.; Sander, C.; Simons, K.; Zerial, M.  
 Mol. Cell. Biol. 10, 6578-6585, 1990  
 A:Title: Molecular cloning of YP1/SEC4-related CDNA's from an epithelial cell line.  
 A:Reference number: A36364; MUID:91061765; PMID:2123294  
 A:Accession: A36364  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-157, 'EK', 160-205 <CHA>  
 A:Cross-references: GB:X56384; NID:9913  
 C:Function:  
 A:Description: probably involved in protein transport from the endoplasmic reticulum thru  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipoprotein; membrane f  
 tein  
 F:12-127/Domain: translation elongation factor Tu homology <ENT>  
 F:18-25/Region: nucleotide-binding motif A (P-loop)  
 F:62-67/Region: nucleotide-binding motif B  
 F:124-127/Region: GTP-binding NKXD motif  
 F:154-156/Region: GTP-binding SAK/L motif  
 F:24,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
 F:194/Binding site: phosphate (Ser) (covalent) (by cdcc kinase) #status predicted  
 F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 97.1%; Score 860; DB 1; Length 205;  
 Best Local Similarity 84.4%; Pred. No. 2.2e-62;  
 Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;  
 QY 1 MSSNPEYDLFKLLIDSGVSKCLLRPADDTYTSYSTIGVDKIRIETLDGKI 60  
 |||||  
 Db 1 MSSNPEYDLFKLLIDSGVSKCLLRPADDTYTSYSTIGVDKIRIETLDGKI 60  
 QY 61 KIQI-----ESFNNYKQWLOEIDRYASENVKL 88  
 |||||  
 Db 61 KQIWDTAGQERFRTITSSYRGAGIIVYDVDTQESFNNYKQWLOEIDRYASENVKL 120  
 QY 89 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKKRMGPGA 148  
 Db 121 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKKRMGPGA 180  
 QY 149 TAGGAEKSNVKIQSTPVRKSGGCC 173  
 |||||  
 Db 181 TAGGAEKSNVKIQSTPVRKSGGCC 205

## RESULT 3

TYMSYP

GTP-binding protein ypt1 - mouse  
 N:Alternate names: GTP-binding protein Rab1; ras-related protein ypt1; transforming p  
 C:Species: Mus musculus (house mouse)  
 C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 19-Jan-2001  
 C:Accession: S05551; S06285  
 R:Wichmann, H.; Disela, C.; Haubruck, H.; Gallwitz, D.  
 Nucleic Acids Res. 17, 6737-6738, 1989  
 A:Title: Nucleotide sequence of the mouse ypt1 gene encoding a ras-related GTP-bindin  
 A:Reference number: S05551; MUID:89386011; PMID:2506528  
 A:Accession: S05551  
 A:Molecule type: DNA  
 A:Residues: 1-205 <MIC>  
 A:Cross-references: EMBL:X15744; NID:955458; PIDN:CAA33760.1; PID:9763158  
 R:Haubruck, H.; Disela, C.; Wagner, P.; Gallwitz, D.  
 EMBO J. 6, 4049-4053, 1987  
 A:Title: The ras-related ypt protein is an ubiquitous eukaryotic protein: isolation a  
 A:Reference number: S06285; MUID:88166649; PMID:3127202  
 A:Accession: S06285  
 A:Molecule type: mRNA  
 A:Residues: 1-205 <HAU>  
 C:Genetics:  
 A:Gene: ypt1  
 A:Introns: 8/2; 32/3; 64/3; 96/3; 140/3  
 C:Function:  
 A:Description: probably involved in protein transport from the endoplasmic reticulum  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipoprotein; membran  
 tein  
 F:12-127/Domain: translation elongation factor Tu homology <ENT>  
 F:18-25/Region: nucleotide-binding motif A (P-loop)  
 F:62-67/Region: nucleotide-binding motif B  
 F:124-127/Region: GTP-binding NKXD motif  
 F:154-156/Region: GTP-binding SAK/L motif  
 F:24,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #  
 F:194/Binding site: phosphate (Ser) (covalent) (by cdcc kinase) #status predicted  
 F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 97.1%; Score 860; DB 1; Length 205;  
 Best Local Similarity 84.4%; Pred. No. 2.2e-62;  
 Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;  
 QY 1 MSSNPEYDLFKLLIDSGVSKCLLRPADDTYTSYSTIGVDKIRIETLDGKI 60  
 |||||  
 Db 1 MSSNPEYDLFKLLIDSGVSKCLLRPADDTYTSYSTIGVDKIRIETLDGKI 60  
 QY 61 KIQI-----ESFNNYKQWLOEIDRYASENVKL 88  
 |||||  
 Db 61 KQIWDTAGQERFRTITSSYRGAGIIVYDVDTQESFNNYKQWLOEIDRYASENVKL 120  
 QY 89 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKKRMGPGA 148  
 Db 121 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKKRMGPGA 180  
 QY 149 TAGGAEKSNVKIQSTPVRKSGGCC 173  
 |||||  
 Db 181 TAGGAEKSNVKIQSTPVRKSGGCC 205

## RESULT 4

TYRTYP

GTP-binding protein Rab1 - rat  
 N:Alternate names: transforming protein ypt1 homology  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 03-Aug-1992 #sequence\_revision 06-Dec-1996 #text\_change 19-Jan-2001  
 C:Accession: A39963  
 R:Touche, N.; Chardin, P.; Tavittian, A.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 8210-8214, 1987  
 A:Title: Four additional members of the ras gene superfamily isolated by an oligonuci  
 A:Reference number: A39963; MUID:88068563; PMID:3317403  
 A:Accession: A39963  
 A:Molecule type: mRNA  
 A:Residues: 1-205 <TOU>  
 A:Cross-references: GB:J02998; NID:9206552; PIDN:AAAA2006.1; PID:9206553



C:Function: probably involved in protein transport from the endoplasmic reticulum th  
 A:Description: ras transforming protein; translation elongation factor Tu homology  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipoprotein; membrane  
 F:12-127/Domain: translation elongation factor Tu homology <ETU>  
 F:18-25/Region: nucleotide-binding motif A (P-loop)  
 F:62-67/Region: nucleotide-binding motif B  
 F:124-127/Region: GTP-binding NKXD motif  
 F:154-156/Region: GTP-binding SAK/L motif  
 F:24,25,43,124,125,127,154/Binding site: Mg-GTP (lys, Ser, Thr, Asp, Lys, Asp, Ser) #sta  
 F:194/Binding site: phosphate (Ser) (covalent) (by cdcc kinase) #status predicted  
 F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 95.8%; Score 849; DB 1; Length 205;  
 Best Local Similarity 83.4%; Pred. No. 1,7e-61;  
 Matches 171; Conservative 0; Mismatches 2; Indels 32; Gaps 1;

OY 1 MSNPEYDYLKLLIGDSGVKSCILLRFADDTYESISTIGVDFKIRITELDGKTI 60  
 DB 1 MSNPEYDYLKLLIGDSGVKSCILLRFADDTYESISTIGVDFKIRITELDGKTI 60  
 OY 61 KIQI-----ESFNNVKQWLQETIDRYASENVNKL 88  
 DB 61 KIQIDTAGQERFRTTSSYRGAGIIIVYDVTDOESFNNVKQWLQETIDRYASENVNKL 120  
 OY 89 LVGNKCDLTTKKVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTMAEIKRKGPGA 148  
 DB 121 LVGNKCDLTTKKVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTMAEIKRKGPGA 180  
 OY 149 TAGGAESNVKIOSTPVKSGGCGC 173  
 DB 181 TAGGAESNVKIOSTPVKSGGCGC 205

RESULT 5  
 D38625  
 C:Function: GTP-binding protein o-rab1 - electric ray (Discopyge ommata)  
 C:Species: Discopyge ommata  
 C:Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 02-Feb-2001  
 C:Accession: D38625  
 R:Nggee, J.K.; Elferink, L.A.; Scheller, R.H.  
 J. Biol. Chem. 266, 2675-2680, 1991  
 A:Title: A family of ras-like GTP-binding proteins expressed in electrorotor neurons.  
 A:Reference number: A38625; MUID:91115900; PMID:1899244  
 A:Accession: D38625  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-201 <NGS>  
 A:Cross-references: GB:M38393  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKXD motif  
 F:150-152/Region: GTP-binding SAK/L motif  
 F:21,22,40,121,122,124,150/Binding site: Mg-GTP (lys, Ser, Thr, Asp, Lys, Asp, Ser) #sta  
 F:200,201/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 91.7%; Score 812.5; DB 2; Length 201;  
 Best Local Similarity 81.2%; Pred. No. 1.5e-58;  
 Matches 164; Conservative 3; Mismatches 2; Indels 33; Gaps 2;

OY 4 MNPEDYDLKLLIGDSGVKSCILLRFADDTYESISTIGVDFKIRITELDGKTIKQ 63  
 DB 1 MNPEDYDLKLLIGDSGVKSCILLRFADDTYESISTIGVDFKIRITELDGKTIKQ 60  
 OY 64 I-----ESFNNVKQWLQETIDRYASENVNKL 91  
 DB 61 IWDTAGQERFRTTSSYRGAGIIIVYDVTDOESFNNVKQWLQETIDRYASENVNKL 120  
 OY 92 NKCDLTTKKVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTMAEIKRKGPGA 151  
 DB 121 NKCDLTTKKVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTMAEIKRKGPGA 180  
 OY 152 GAESNVKIOSTPVKSGGCGC 173  
 DB 181 G-ERPMLKIDSTPVKSGGCGC 201

DB 121 NKCDLTTKKVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTMAEIKRKGPGA 179  
 OY 152 GAESNVKIOSTPVKSGGCGC 173  
 DB 180 GSEKSNVNIQSTPVKSGGCGC 201

RESULT 6  
 S06147  
 C:Function: GTP-binding protein rab1 - rat  
 N:Alternate names: ras-related protein rab1  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 02-Feb-2001  
 C:Accession: S06147; S03189  
 R:Touchot, N.; Zahraoui, A.; Vielh, E.; Tavilian, A.  
 FEBS Lett. 236, 79-84, 1989  
 A:Title: Biochemical properties of the YPT-related rab1 protein. Comparison with rab  
 A:Reference number: S06147; MUID:90033316; PMID:2509243  
 A:Accession: S06147  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-201 <TUD>  
 R:Touchot, N.; Zahraoui, A.; Vielh, E.; Tavilian, A.  
 Nucleic Acids Res. 17, 1770, 1989  
 A:Title: Nucleotide sequence of a rat cDNA: RAB1B, encoding a RAB1-YPT related protel  
 A:Reference number: S03189; MUID:89160341; PMID:2493636  
 A:Accession: S03189  
 A:Molecule type: mRNA  
 A:Residues: 1-95, 'YA', 98-201 <VIE>  
 A:Cross-references: EMBL:X13905; MUID:957005; PIDN:CAA32105.1; PID:957006  
 C:Genetics:  
 A:Gene: rab1B  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKXD motif  
 F:151-153/Region: GTP-binding SAK/L motif  
 F:21,22,40,121,122,124,151/Binding site: Mg-GTP (lys, Ser, Thr, Asp, Lys, Asp, Ser) #  
 F:200,201/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 87.1%; Score 771.5; DB 2; Length 201;  
 Best Local Similarity 77.2%; Pred. No. 3e-55;  
 Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

OY 4 MNPEDYDLKLLIGDSGVKSCILLRFADDTYESISTIGVDFKIRITELDGKTIKQ 63  
 DB 1 MNPEDYDLKLLIGDSGVKSCILLRFADDTYESISTIGVDFKIRITELDGKTIKQ 60  
 OY 64 I-----ESFNNVKQWLQETIDRYASENVNKL 91  
 DB 61 IWDTAGQERFRTTSSYRGAGIIIVYDVTDOESFNNVKQWLQETIDRYASENVNKL 120  
 OY 92 NKCDLTTKKVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTMAEIKRKGPGA 151  
 DB 121 NKCDLTTKKVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTMAEIKRKGPGA 180  
 OY 152 GAESNVKIOSTPVKSGGCGC 173  
 DB 181 G-ERPMLKIDSTPVKSGGCGC 201

RESULT 7  
 S38339  
 C:Function: GTP-binding protein rab1 - great pond snail  
 C:Species: Lymnaea stagnalis (great pond snail)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 02-Feb-2001  
 C:Accession: S38339; S32206  
 R:Alteberg, M.; van Die, I.; Yang, H.; Andriessen, J.A.; van Tetering, A.; van den E  
 Eur. J. Biochem. 217, 241-246, 1993  
 A:Title: Isolation and characterization of three cDNAs coding for Rab proteins from t  
 A:Reference number: S38339; MUID:94039042; PMID:8223561  
 A:Accession: S38339

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-205 <AGT>  
A:Cross-references: EMBL:X72688; NID:g288933; PIDN:CAA51233.1; PID:g288934  
C:Genetics:  
A:Gene: rab1  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; 11poprotein; nucleotide binding; P-loop; prenylated cysteine  
F:12-127/Domain: translation elongation factor Tu homology <ETD>  
F:18-25/Region: nucleotide-binding motif A (P-loop)  
F:124-127/Region: GTP-binding NXXD motif  
F:154-156/Region: GTP-binding SAR/L motif  
F:24,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #str  
F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 82.7%; Score 733; DB 2; Length 205;  
Best Local Similarity 73.8%; Pred. No. 4e-52;  
Matches 152; Conservative 7; Mismatches 13; Indels 34; Gaps 3;

QY 1 MSANPEYDLKLLIGSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDGKTI 60  
DB 1 MSTMPDYDLKLLIGSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDGKTI 60  
QY 61 KLOI-----ESFNWKQWLOEIDRYASENVNKL 88  
DB 61 KLOIMDTAGQERFRITTSYRGAGIIVYDVTDOESFNWKQWLOEIDRYASENVNKL 120  
QY 89 LVGNKCDLTKKRVYDTTAKERFADSLGIFPLETSANANVQSEFMTAAEIKRRGPGA 148  
DB 121 LVGNKSDLTTRKRVDPDTAKERADQIGIFPLETSANANVQSEFMTAAEIKRRGPG-1 179  
QY 149 TAGGAESNVKI-QSTPVKQSGGCC 173  
DB 180 TAASDPSKPSYKINSSTPVSAKGGCC 205

RESULT 8  
JEO318  
GTP-binding protein rabB - silkworm  
C:Species: Bombyx mori (silkworm)  
C:Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 02-Feb-2001  
C:Accession: JEO318  
R:Uno, T.; Uno, M.; Nakajima, A.; Shirai, Y.; Alconzo, Y.  
Biosci. Biotechnol. Biochem. 62, 1885-1891, 1998  
A:Title: Molecular cloning of cDNA for Rab from the brain of Bombyx mori and biochemical  
A:Reference number: JEO318; M0ID:99053143; PMID:9836423  
A:Accession: JEO318  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-202 <UNO>  
A:Cross-references: GB:AF013572  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:9-124/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:121-124/Region: GTP-binding NXXD motif  
F:151-153/Region: GTP-binding SAR/L motif  
F:21,22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

Query Match 75.7%; Score 671; DB 2; Length 202;  
Best Local Similarity 66.5%; Pred. No. 4e-47;  
Matches 135; Conservative 18; Mismatches 16; Indels 34; Gaps 3;

QY 4 NMPEDYELKLLIGSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDGKTIKIQ 63  
DB 1 NMPEDYELKLLIGSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDGKTIKIQ 60  
QY 64 I-----ESFNWKQWLOEIDRYASENVNKL 91  
DB 61 IMDTAGQERFRITTSYRGAGIIVYDCTDSDSNWKQWLOEIDRYASENVNKL 120  
QY 92 NKCIDLTKKRVYDTTAKERFADSLGIFPLETSANANVQSEFMTAAEIKRRGPGA 151  
DB 92 NKCIDLTKKRVYDTTAKERFADSLGIFPLETSANANVQSEFMTAAEIKRRGPGA 151

DB 121 NKCIDLTKKRVYDTTAKERFADSLGIFPLETSANANVQSEFMTAAEIKRRGPGA 179  
QY 152 GAESNVKI-QSTPVKQSGGCC 173  
DB 180 MAPAGHKIDQGPIDTGKSSCC 202

RESULT 9  
T33781  
hypothetical protein C39F7.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-Jan-2000  
C:Accession: T33781  
R:Magill, L.; Scheet, P.; Dubbelde, C.  
submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of C. elegans cosmid C39F7.  
A:Reference number: Z21407  
A:Accession: T33781  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-205 <MAG>  
A:Cross-references: EMBL:AF101310; PIDN:AC69218.1; GSPDB:GN00023; CESP:C39F7.4  
A:Experimental source: strain Bristol N2; clone C39F7  
C:Genetics:  
A:Gene: CESP:C39F7.4  
A:Map position: 5  
A:Insertions: 8/2; 48/3; 64/3; 96/3; 163/3  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 74.5%; Score 660; DB 2; Length 205;  
Best Local Similarity 66.2%; Pred. No. 3.2e-46;  
Matches 137; Conservative 16; Mismatches 18; Indels 36; Gaps 4;

QY 1 MSANPEYDLKLLIGSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDGKTI 60  
DB 1 MAANPEYDLKLLIGSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDGKTI 60  
QY 61 KLOI-----ESFNWKQWLOEIDRYASENVNKL 88  
DB 61 KLOIMDTAGQERFRITTSYRGAGIIVYDVTDOETFNWKQWLOEIDRYASENVNKL 120  
QY 89 LVGNKCDLTKKRVYDTTAKERFADSLGIFPLETSANANVQSEFMTAAEIKRRGPGA 148  
DB 121 LVGNKCDLTKKRVYDTTAKERFADSLGIFPLETSANANVQSEFMTAAEIKRRGPGA 180  
QY 149 TAGGAESNVKI-QSTPVKQSGGCC 173  
DB 181 GAGGA--PEVRLTGSGPVDKRSKGGCC 205

RESULT 10  
S30096  
GTP-binding protein ypt1 [similarity] - Neurospora crassa  
N:Alternate names: protein B9J10.240; ras-related protein ypt1  
C:Species: Neurospora crassa  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Jan-2001  
C:Accession: S30096; T49787  
R:Rehltz, K.; Palme, K.; Dieffenhal, T.; Russo, V.E.A.  
Mol. Gen. Genet. 235, 413-421, 1992  
A:Title: The Ncyp1 gene from Neurospora crassa is located on chromosome 2: molecular  
A:Reference number: S30096; M0ID:93101148; PMID:1361212  
A:Accession: S30096  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-203 <HEI>  
A:Cross-references: GB:S51252; NID:g262045; PIDN:AA824564.1; PID:g262046  
R:Schulte, U.; Align, V.; Hohnel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatu  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49787  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-203 <SCH>

A:Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.240  
 A:Experimental source: BAC clone B9J10; strain OR74A  
 C:Genetics:  
 A:Gene: ypt1; NCSP:B9J10.240  
 A:Map position: 6  
 A:Introns: 4/2; 45/3; 61/3; 137/3  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; lipoprotein; nucleotide binding; P-loop; prenylated cysteine  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKX motif  
 F:151-153/Region: GTP-binding NKX motif  
 F:21-22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
 F:202,203/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 74.3%; Score 658.5; DB 2; Length 203;  
 Best Local Similarity 66.5%; Pred. No. 4,2e-46;  
 Matches 135; Conservative 11; Mismatches 24; Indels 33; Gaps 2;

QY 4 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDKRTIKLQ 63  
 |||||  
 DB 1 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDKRTIKLQ 60

QY 64 I-----ESFNNVKOMLQIEDRYASENVKLLVG 91  
 |||||  
 DB 61 IMPTAGGERFRTITSSYRGAGHICVYVDYTDMSFNNVKOMLQIEDRYATEGVNKLIVG 120

QY 92 NKCDLTKKVVDTYTAKEFADSLGIPLETSAKNATNVESFMTMAEIKRKMGATAG 151  
 |||||  
 DB 121 NKSDMDKRVYEVYAKAFADSLGIPLETSAKNASNVQAFITMAROIKRMGSSTATN 180

QY 152 GAERKSNVKI-QSTPYKOSGGGCC 173  
 |||||  
 DB 181 NTKASVNVSRGCVSNSSGCC 203

RESULT 11  
 JC4105  
 GTP-binding protein yptc1 - Chlamydomonas reinhardtii  
 N:Alternate names: membrane vesicle transport protein yptc1; ras-like yptc1 protein; smc  
 C:Species: Chlamydomonas reinhardtii  
 C:Date: 21-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 02-Feb-2001  
 C:Accession: JC4105  
 R:Dietmaier, W.; Fabry, S.; Huber, H.; Schmitt, R.  
 Gene 158, 41-50, 1995  
 A:title: Analysis of a family of ypt genes and their products from Chlamydomonas reinhar  
 A:Reference number: JC4105; MUID:95309723; PMID:7789809  
 A:Accession: JC4105  
 A:Molecule type: mRNA  
 A:Residues: 1-203 <ID>  
 A:Cross-references: GB:U13168; NID:9806723; PID:AAA82727.1; PID:9806724  
 C:Comment: This protein plays an essential role in the regulation of intracellular memb  
 C:Genetics:  
 A:Gene: yptc1  
 A:Introns: 5/1; 10/2; 29/2; 45/2; 61/2; 85/2; 137/2; 172/1  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKX motif  
 F:151-153/Region: GTP-binding NKX motif  
 F:21-22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
 F:202,203/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 73.0%; Score 646.5; DB 2; Length 203;  
 Best Local Similarity 65.7%; Pred. No. 3.9e-45;  
 Matches 134; Conservative 9; Mismatches 26; Indels 35; Gaps 3;

QY 4 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDKRTIKLQ 63  
 |||||  
 DB 1 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDKRTIKLQ 60

QY 64 I-----ESFNNVKOMLQIEDRYASENVKLLVG 91  
 |||||  
 DB 61 IMPTAGGERFRTITSSYRGAGHICVYVDYTDMSFNNVKOMLQIEDRYASENVKLLVG 120

QY 92 NKCDLTKKVVDTYTAKEFADSLGIPLETSAKNATNVESFMTMAEIKRKMG--PGAT 149  
 |||||  
 DB 121 NKSDMDKRVYEVYAKAFADSLGIPLETSAKNATNVQDAFMTMAEIKRMASQPIPT 180

QY 150 AGAERKSNVKI-QSTPYKOSGGGCC 173  
 |||||  
 DB 181 KAGGPVVRQ-EGKPIKSSGCC 203

RESULT 12  
 T50323  
 ypt1-related protein 1 [imported] - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 02-Sep-2000  
 C:Accession: T50323  
 R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaure, V.; Gilbert  
 submitted to the EMBL Data Library, January 2000  
 A:Reference number: 225061  
 A:Accession: T50323  
 A:Status: preliminary; translated from GB/EMBL/DDBB  
 A:Molecule type: DNA  
 A:Residues: 1-203 <MCD>  
 A:Cross-references: EMBL:AL356326; PID:CA66454.1; GSPDB:GN00067; SPDB:SPBC1703.10  
 C:Genetics:  
 A:Gene: SPDB:SPBC1703.10  
 A:Map position: 2  
 A:Introns: 4/2; 45/3; 61/3; 94/1  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>

Query Match 73.0%; Score 646.5; DB 2; Length 203;  
 Best Local Similarity 64.5%; Pred. No. 3.9e-45;  
 Matches 131; Conservative 16; Mismatches 23; Indels 33; Gaps 2;

QY 4 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDKRTIKLQ 63  
 |||||  
 DB 1 MNPEYDLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDKRTIKLQ 60

QY 64 I-----ESFNNVKOMLQIEDRYASENVKLLVG 91  
 |||||  
 DB 61 IMPTAGGERFRTITSSYRGAGHICVYVDYTDMSFNNVKOMLQIEDRYASENVKLLVG 120

QY 92 NKCDLTKKVVDTYTAKEFADSLGIPLETSAKNATNVESFMTMAEIKRKMGATAG 151  
 |||||  
 DB 121 NKSDMDKRVYEVYAKAFADSLGIPLETSAKNATNVQDAFMTMAEIKRMGSSTATN 180

QY 152 GAERKSNVKI-QSTPYKOSGGGCC 173  
 |||||  
 DB 181 NTKASVNVSRGCVSNSSGCC 203

RESULT 13  
 JC1247  
 GTP-binding protein yptv1 - Volvox carterii  
 C:Species: Volvox carterii  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 02-Feb-2001  
 C:Accession: JC1247  
 R:Fabry, S.; Natr, N.; Huber, H.; Palme, K.; Jaenicke, L.; Schmitt, R.  
 Gene 116, 153-162, 1992  
 A:title: The yptv1 gene encodes a small G-protein in the green alga Volvox carterii: G  
 A:Reference number: JC1247; MUID:92280499; PMID:1511889  
 A:Accession: JC1247  
 A:Molecule type: DNA  
 A:Residues: 1-203 <FAB>  
 A:Cross-references: GB:M3438; NID:9170660; PID:AAA4255.1; PID:9170661  
 C:Genetics:  
 A:Gene: yptv1  
 A:Introns: 5/2; 10/3; 29/3; 45/3; 61/3; 85/3; 137/3; 172/2



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 15:15:45 ; Search time 22 Seconds  
(without alignments)  
326.155 Million cell updates/sec

Title: US-09-820-003b-2

Profile score: 886  
Sequence: 1 MSSWNPEDYDLFKLLIGDS.....EKSWNKIQSTPVMQSGGCGC 173

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860	97.1	205	1	RB1A_HUMAN
2	849	95.8	205	1	RB1A_RAT
3	828	93.5	202	1	RB1L_DISOM
4	760.5	85.8	201	1	RB1B_RAT
5	733	82.7	205	1	RB1L_TYMS
6	658.5	74.3	203	1	YPT1_MENCR
7	646.5	73.0	203	1	YPT1_CHURE
8	646.5	73.0	203	1	YPT1_SCHPO
9	642.5	72.5	203	1	YPT1_VOICA
10	631.5	71.3	201	1	YPT1_PHTIN
11	619.5	69.3	208	1	ARAS_ARATH
12	618.5	69.2	203	1	YPT2_MAIZE
13	613.5	69.2	202	1	RIC1_ORISA
14	571	64.4	206	1	YPT1_YEAST
15	542	61.2	167	1	RB1A_DICDI
16	538	60.7	199	1	RB1B_DICDI
17	523	59.0	208	1	YPT1_MAIZE
18	468.5	52.9	203	1	SAS2_DICDI
19	461	52.0	203	1	SAS2_DICDI
20	438.5	49.9	215	1	RB1L_BETVU
21	432	49.5	216	1	ARAS_ARATH
22	435.5	49.2	200	1	YPT2_SCHPO
23	424	47.9	210	1	RB1B_DISOM
24	418	47.2	207	1	RB1B_HUMAN
25	417	47.1	207	1	RB1B_RAT
26	414.5	46.8	199	1	RB1A_DICDI
27	410.5	46.7	207	1	RB1B_HUMAN
28	410.5	46.3	206	1	RB1B_MOUSE
29	405.5	45.8	217	1	YPT2_VOICA
30	403.5	45.5	200	1	RB1L_HUMAN
31	402.5	45.4	200	1	RB1L_CANFA
32	398	44.9	200	1	RB1L_DISOM
33	398	44.9	203	1	RYL1_YARLI

34	384.5	43.4	203	1	RB13_HUMAN
35	374.5	42.3	200	1	RB1O_RAT
36	373.5	42.2	215	1	SEC4_YEAST
37	360.5	40.7	201	1	RB35_HUMAN
38	360	40.6	210	1	SEC4_CANAL
39	352	39.7	209	1	RB2A_MAIZE
40	349.5	39.4	210	1	RB2B_MAIZE
41	344.5	38.9	213	1	YPT4_CHURE
42	342	38.6	219	1	RB3A_BOVIN
43	339	38.3	213	1	YPT4_VOICA
44	336	37.9	220	1	RB3B_DROME
45	336	37.9	220	1	RB3A_HUMAN

## ALIGNMENTS

RESULT 1	RB1A_HUMAN	STANDARD;	PRT;	205 AA.
AC	P1476:			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Ras-related protein Rab-1A (YPT1-related protein).			
GN	RB1 OR RB1A.			
OS	Homo sapiens (Human), and			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID:9606, 10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human;			
RA	MEDLINE=89308668; PubMed=2501306;			
RA	Zahraoui A., Touchot N., Charlin P., Tavilian A.;			
RT	"The human Rab genes encode a family of GTP-binding proteins related			
RT	to yeast YPT1 and SEC4 products involved in secretion.";			
RL	J. Biol. Chem. 264:12394-12401(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human; TISSUE=Placenta;			
RA	Strausberg R.;			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Mouse;			
RA	MEDLINE=8816649; PubMed=3127202;			
RA	Haudruck H., Disela C., Wagner P., Gallwitz D.;			
RT	"The ras-related ypt protein is an ubiquitous eukaryotic protein:			
RT	isolation and sequence analysis of mouse cDNA clones highly			
RT	homologous to the yeast YPT1 gene.";			
RL	EMBO J. 6:4049-4053(1987).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Mouse;			
RA	MEDLINE=89386011; PubMed=2506528;			
RA	Wichmann H., Disela C., Haudruck H., Gallwitz D.;			
RT	"Nucleotide sequence of the mouse ypt1 gene encoding a ras-related			
RT	GTP-binding protein.";			
RT	Nucleic Acids Res. 17:6737-6738(1989).			
RL	[5]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=Human;			
RA	Wu G., Dorn G.W. II;			
RT	"Mouse Rab1a, member of Ras oncogene family mRNA.";			
RT	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	PHOSPHORYLATION BY CDC2.			
RA	MEDLINE=9121885; PubMed=1902553;			
RA	Bailly E., McCreffey M., Touchot N., Zahraoui A., Goud B., Bornens M.;			
RT	"Phosphorylation of two small GTP-binding proteins of the Rab family			
RT	by p34cdc2.";			

RL Nature 350:715-718(1991).  
 RN [7]  
 RP ISOPRENOLD.  
 RX MEDLINE-91296801; PubMed-1648736;  
 RA Khosravi-Far R., Lutz R.J., Cox A.D., Conroy L., Bourne J.R.,  
 RT Shinsky M., Balch W.E., Buss J.E., Der C.J.,  
 "Isoprenoid modification of rab proteins terminating in CC or CXC  
 motifs".  
 CC Proc. Natl. Acad. Sci. U.S.A. 88:6264-6268(1991).  
 CC -1- FUNCTION: PROBABLY REQUIRED FOR TRANSIT OF PROTEIN FROM THE ER  
 CC THROUGH GOLGI COMPARTMENT.  
 CC -1- SUBCELLULAR LOCATION: GOLGI COMPARTMENT.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2-KINASE DURING MITOSIS.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M28209; AAA60240.1; -  
 DR EMBL: BC000905; AAH00905.1; -  
 DR EMBL: Y00094; CAA68284.1; -  
 DR EMBL: X15744; CAA33760.1; -  
 DR EMBL: X15745; CAA33760.1; JOINED.  
 DR EMBL: X15746; CAA33760.1; JOINED.  
 DR EMBL: X15747; CAA33760.1; JOINED.  
 DR EMBL: AF226873; AA633844.1; -  
 DR PIR: S05551; TWSTP.  
 DR PIR: A34323; A34323.  
 DR PIR: S19104; S19104.  
 DR HSSP: P05713; 3RAB.  
 DR PMMA-2DPAGE: P11476; -  
 DR GeneW: HGNC:9758; RAB1A.  
 DR MIM: 179508; -  
 DR MGD: MGI:97842; Rab1.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAAS: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack;  
 KW Endoplasmic reticulum; Phosphorylation.  
 FT NP\_BIND 18 25 GTP (BY SIMILARITY).  
 FT NP\_BIND 66 70 GTP (BY SIMILARITY).  
 FT NP\_BIND 124 127 GTP (BY SIMILARITY).  
 FT DOMAIN 40 48 EFFECTOR REGION (BY SIMILARITY).  
 FT MOD\_RES 194 194 PHOSPHORYLATION (BY CDC2) (PROBABLE).  
 FT LIPID 204 204 GERANYL-GERANYL.  
 FT LIPID 205 205 GERANYL-GERANYL.  
 SQ SEQUENCE 205 AA: 22678 MW: 22678 MW: B2A8F43B0FB17D6 CRC64:  
 Query Match 97.1%; Score 860; DB 1; Length 205;  
 Best Local Similarity 84.4%; Pred. No. 9, 3e-65;  
 Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 149 TAGAERSNVKIOSTPVKSGGCC 173  
 DB 181 TAGAERSNVKIOSTPVKSGGCC 205  
 RESULT 2  
 ID RBLA\_RAT STANDARD; PRT; 205 AA.  
 AC P05711;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ras-related protein Rab-1A.  
 GN RAB1 OR RAB1A.  
 OS Rattus norvegicus (Rat), and  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116, 9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Rat.  
 RX MEDLINE-88068563; PubMed-3317403;  
 RA Touchot N., Chardin P., Tavilian A.;  
 RT Four additional members of the ras gene superfamily isolated by an  
 RT oligonucleotide strategy: molecular cloning of YPT-related cDNAs from  
 RT a rat brain library".  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8210-8214(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-C.familiaris; STRAIN-Cocker spaniel;  
 RX MEDLINE-91061765; PubMed-2133294;  
 RA Chavrier P., Vangron M., Sander C., Simons K., Zerial M.;  
 RT Molecular cloning of YP11/SEC4-related cDNAs from an epithelial cell  
 RT line.".   
 RL Mol. Cell. Biol. 10:6578-6585(1990).  
 RN [3]  
 RP ISOPRENOLD.  
 RX MEDLINE-91296801; PubMed-1648736;  
 RA Khosravi-Far R., Lutz R.J., Cox A.D., Conroy L., Bourne J.R.,  
 RT Shinsky M., Balch W.E., Buss J.E., Der C.J.,  
 RT "Isoprenoid modification of rab proteins terminating in CC or CXC  
 RT motifs".  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6264-6268(1991).  
 CC -1- FUNCTION: PROBABLY REQUIRED FOR TRANSIT OF PROTEIN FROM THE ER  
 CC THROUGH GOLGI COMPARTMENT.  
 CC -1- MISCELLANEOUS: RAB-1A BINDS GTP AND GDP AND POSSESS INTRINSIC  
 CC GTPASE ACTIVITY.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J02998; AAA42006.1; -  
 DR EMBL: X56384; CAB56775.1; ALT\_SEQ.  
 DR PIR: A36364; A36364.  
 DR PIR: A39963; A39963.  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAAS: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack;  
 KW Endoplasmic reticulum.  
 FT NP\_BIND 18 25 GTP (BY SIMILARITY).

```

FT NP_BIND 66 70 GTP (BY SIMILARITY).
FT NP_BIND 124 127 GTP (BY SIMILARITY).
FT DOMAIN 40 48 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 204 204 GERANYL-GERANYL.
FT LIPID 205 205 GERANYL-GERANYL.
SQ SEQUENCE 205 AA; 22763 MW; 34649523B0FB0E6FC CRC64;

Query Match 95.8%; Score 849; DB 1; Length 205;
Best Local Similarity 83.4%; Pred. No. 7.7e-64;
Matches 171; Conservative 0; Mismatches 2; Indels 32; Gaps 1;

QY 1 MSNPNPYDLFKLLIGDSGVGKSCLLRFADPTYESYISITIGVDFKRIEIDGKTI 60
DB 1 MSNPNPYDLFKLLIGDSGVGKSCLLRFADPTYESYISITIGVDFKRIEIDGKTI 60
QY 61 KLOI-----ESFNNVKOMLOEIDRYASENVKL 88
DB 61 KLOIIMDTAGGERFRTTSSYRGAGIIVYDVTDQESFNNVKOMLOEIDRYASENVKL 120
QY 89 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATVROSEFMTMAEIKRMGPGA 148
DB 121 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATVROSEFMTMAEIKRMGPGA 180
QY 149 TAGGAESKNVKIOSPTPVKSGGCC 173
DB 181 TAGGAESKNVKIOSPTPVKSGGCC 205

RESULT 3
RAB1_DISOM
ID RAB1_DISOM STANDARD: PRT; 202 AA.
AC P2125;
DR 01-AUG-1991 (Rel. 19, Created)
DR 01-AUG-1991 (Rel. 19, Last sequence update)
DR 01-OCT-1994 (Rel. 30, Last annotation update)
DE Ras-related protein OKAB-1.
OS Discopylea ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalea; Hynostomata; Pristigaster; Batoidae;
OC OCtopediniiformes; Narcinoidae; Narcinidae; Discopylea.
OX NCBI_Taxid=7785;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Electric lobe;
RA MEDLINE-9115900; PubMed-1899244;
RA Ngeese J.K., Elfertink L.A., Scheller R.H.;
RT "A family of ras-like GTP-binding proteins expressed in electrosensor
neurons."
RL J. Biol. Chem. 266:2675-2680(1991).
CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
TRAFFIC (BY SIMILARITY).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sdb.ch/announce/
or send an email to license@sdb-sdb.ch)
CC -----
EMBL: M38393; AAA4234.1; -
DR PIR: D38625; D38625.
DR HSP: P05713; 3888.
DR InterPro: IPR003578; GTPase_Rab.
DR InterPro: IPR001806; Ras_transfmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; Ras_1.
DR PRINTS: PR00449; RASTNSPRNG.
DR SMART: SM00175; Ras_1.
DR TIGRfams: TIGR00231; small_GTP_1.
DR GTP_Binding: Lipoprotein: Prenylation; Protein transport.
FT NP_BIND 15 22 GTP (BY SIMILARITY).

```

```

FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 202 AA; 22333 MW; 31078502BCDD1B9 CRC64;

Query Match 93.5%; Score 828; DB 1; Length 202;
Best Local Similarity 81.7%; Pred. No. 4.2e-62;
Matches 165; Conservative 3; Mismatches 2; Indels 32; Gaps 1;

QY 4 MNPEYDLFKLLIGDSGVGKSCLLRFADPTYESYISITIGVDFKRIEIDGKTI 63
DB 1 MNPEYDLFKLLIGDSGVGKSCLLRFADPTYESYISITIGVDFKRIEIDGKTI 60
QY 64 I-----ESFNNVKOMLOEIDRYASENVKL 91
DB 61 IMPTAGGERFRTTSSYRGAGIIVYDVTDQESFNNVKOMLOEIDRYASENVKL 120
QY 92 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATVROSEFMTMAEIKRMGPGA 151
DB 121 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATVROSEFMTMAEIKRMGPGA 180
QY 152 GAKSNVKIOSPTPVKSGGCC 173
DB 181 GAKSNVKIOSPTPVKSGGCC 202

RESULT 4
RAB1_RAT
ID RAB1_RAT STANDARD: PRT; 201 AA.
AC P10536;
DR 01-JUL-1989 (Rel. 11, Created)
DR 01-JUL-1989 (Rel. 11, Last sequence update)
DR 01-OCT-1994 (Rel. 30, Last annotation update)
DE Ras-related protein Rab-1b.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-89160341; PubMed-2493636;
RA Zahraoui A., Touchot N., Chardin P., Tavilian A.;
RT "Nucleotide sequence of a rat cDNA: rab1b, encoding a rab1-YPT
related protein."
RL Nucleic Acids Res. 17:1770-1770(1989).
RN [2]
RP CHARACTERIZATION.
RC MEDLINE-90033316; PubMed-2509243;
RA Touchot N., Zahraoui A., Vielh E.
RT "Biochemical properties of the YPT-related rab1b protein. Comparison
with rab1a."
RL FEBS Lett. 256:79-84(1989).
RN [3]
RP ISOPRENOLD.
RC MEDLINE-91296801; PubMed-1648736;
RA Khosravi-Far R., Lutz R.J., Cox A.D., Conroy L., Bourne J.R.,
RA Shensky M., Balch W.E., Buss J.E., Der C.J.;
RT "Isoprenoid modification of rab proteins terminating in CC or CXC
motifs."
RL Proc. Natl. Acad. Sci. U.S.A. 88:6264-6268(1991).
CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
TRAFFIC (BY SIMILARITY).
CC -1- MISCELLANEOUS: RAB-1B BINDS GTP AND GDP AND POSSESS INTRINSIC
GTPASE ACTIVITY.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X13905; CAA32105.1; -  
DR PIR: S03189; S03189.  
DR PIR: S06147; S06147.  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsfmg.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; ras; 1.  
DR PRINTS: PRO0449; RASTRNSFRNG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRfams: TIGR00231; small\_GTP; 1.  
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
FT LIPID 200 200 GERANYL-GERANYL.  
FT LIPID 201 201 GERANYL-GERANYL.  
FT MUTAGEN 21 21 K->M: ABOLISHES GTP-BINDING.  
SQ SEQUENCE 201 AA; 22163 MW; 8D3EEDCA2EFA42FE CRC64;

Query Match 85.8%; Score 760.5; DB 1; Length 201;  
Best Local Similarity 76.2%; Pred. No. 1.7e-56;  
Matches 154; Conservative 7; Mismatches 8; Indels 33; Gaps 2;

QY 4 MNPEVDYLFKLLIGDSGVGSKCLLRADDTYESTYSTIGVDKRTIELDGTI 63  
DB 1 MNPEVDYLFKLLIGDSGVGSKCLLRADDTYESTYSTIGVDKRTIELDGTI 60  
QY 64 I-----ESFNVKQWLOEIDRYASENVKLLVG 91  
DB 61 IMDPAGGERFRVTSYYRGANGIIIVYDVTDOESFNVKQWLOEIDRYASENVKLLVG 120  
QY 92 MKCOLTTKKVVDYTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGANG 151  
DB 121 MKSDLTTRKVVNDYTTAKFADSLGIPLETSAKNATNVEOAFMTAAEIKRMGPGANG 180  
QY 152 GAESKNVKTOSTPVKQSGGCC 173  
DB 181 GERPNLKTIDSTPVKASAGCC 201

RESULT 5  
RAB1\_LYMS  
ID RAB1\_LYMS STANDARD; PRT: 205 AA.  
AC 005974;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE Ras-related protein Rab-1A.  
GN RAB1A.  
OS Lymnaea stagnalis (Great pond snail).  
CC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
CC Lymnaeidae; Lymnaea.  
CC NCBI\_TaxID=6523;  
OX NCBI\_TaxID=6523;  
RN [1]  
RP MEDLINE FROM N.A.  
RX MEDLINE-94039042; PubMed-8223561;  
RA Agterberg M., van Die I., Yang H., Andriessen J.A.,  
van Tetering A., van den Eljnden D.H., Ploegh H.L.;  
\*Isolation and characterization of three cDNAs coding for Rab  
proteins from the alveolar gland of the mollusc Lymnaea stagnalis.\*;  
RT Eur. J. Biochem. 217:241-246(1993).  
RL -1- FUNCTION: PROBABLY REQUIRED FOR TRANSIT OF PROTEIN FROM THE ER  
THROUGH GOLGI COMPARTMENT (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: GOLGI COMPARTMENT (BY SIMILARITY).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X72688; CAA51233.1; -  
DR PIR: S32206; S32206.  
DR PIR: S38339; S38339.  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsfmg.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; ras; 1.  
DR PRINTS: PRO0449; RASTRNSFRNG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRfams: TIGR00231; small\_GTP; 1.  
KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack;  
FT NP\_BIND 18 25 GTP (BY SIMILARITY).  
FT NP\_BIND 66 70 GTP (BY SIMILARITY).  
FT NP\_BIND 124 127 GTP (BY SIMILARITY).  
FT DOMAIN 40 48 EFFECTOR REGION (BY SIMILARITY).  
FT LIPID 204 204 GERANYL-GERANYL (BY SIMILARITY).  
FT LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
SQ SEQUENCE 205 AA; 22760 MW; D5E0BDCB4A26FE CRC64;

Query Match 82.7%; Score 733; DB 1; Length 205;  
Best Local Similarity 73.8%; Pred. No. 3.4e-54;  
Matches 152; Conservative 7; Mismatches 13; Indels 34; Gaps 3;

QY 1 MSSNPEDYLFKLLIGDSGVGSKCLLRADDTYESTYSTIGVDKRTIELDGTI 60  
DB 1 MSTNPEDYLFKLLIGDSGVGSKCLLRADDTYESTYSTIGVDKRTIELDGTI 60  
QY 61 KIQI-----ESFNVKQWLOEIDRYASENVKLL 88  
DB 61 KIQI-----ESFNVKQWLOEIDRYASENVKLL 120  
QY 89 LVGNKCDLTTRKVVNDYTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGANG 148  
DB 121 LVGNKSDLTTRKVVNDYTTAKFADSLGIPLETSAKNATNVEOAFMTAAEIKRMGPGANG 179  
QY 149 TAGGAESKNVKTOSTPVKQSGGCC 173  
DB 180 TAASDKPSVKINSTPVKANKGCC 205

RESULT 6  
YPT1\_NEUCR  
ID YPT1\_NEUCR STANDARD; PRT: 203 AA.  
AC P33723;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE GTP-binding protein ypt1.  
GN YPT-1 OR B9J10.240.  
GN Neurospora crassa.  
OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
CC Sordariales; Sordariaceae; Neurospora.  
CC NCBI\_TaxID=5141;  
OX NCBI\_TaxID=5141;  
RN [1]  
RP MEDLINE FROM N.A.  
RX MEDLINE-93101148; PubMed-1161212;  
RA Helutz K., Palme K., Diefenthal T., Russo V.E.A.;  
\*The Nycp1 gene from Neurospora crassa is located on chromosome 2;  
RT molecular cloning and structural analysis.\*;  
RL Mol. Gen. Genet. 235:413-421(1992).  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-74-OR23-1A;  
RA Schulte U., Aign V., Hohelsel J., Brandt P., Farman B., Holland R.,



RA Miyakeura G., Mewes H.-W., Mannhaupt G.;  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: S51252; AAB24564.1;  
 DR EMBL: AL356324; CAB92031.1;  
 DR PIR: S30096; S30096.  
 DR HSSP: P01112; 1PLT.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsmfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF000071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRMNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfam: TIGR00231; small\_GTP; 1.  
 DR GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 KW NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 203 203 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 203 AA; 22476 MW; 20F51AB3400EF1D CRC64;  
 Query Match 74.3%; Score 658.5; DB 1; Length 203;  
 Best Local Similarity 66.5%; Pred. No. 5.3e-48;  
 Matches 135; Conservative 11; Mismatches 24; Indels 33; Gaps 2;  
 QY 4 MNPEYDYELFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDGKTIKQ 63  
 DB 1 MNPEYDYELFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDGKTIKQ 60  
 QY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91  
 DB 61 IMPYAGQERFRTITSSYRGAHGICVYDVTDMDSFNNVKOMLOEIDRYATEGVNKLIVG 120  
 QY 92 NKCDLTKRKVVDYTTAKAFADSLGIPLETSAKNATNVEGSPFMMAAEIKRKGSCATAG 151  
 DB 121 NKSDMTKAKVVEYTAKEFADSLGIPLETSAKNATNVEGAFETMARQIKEMSGSTAIN 180  
 QY 152 GAERK-NVKIQTSPVKOSGGCC 173  
 DB 181 NTKASVNVSPGCHGVNNSGGCC 203  
 RESULT 7  
 YPFL\_CHLRE STANDARD; PRT; 203 AA.  
 ID YPFL\_CHLRE  
 AC Q39571;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE GTP-binding protein YPFLC1.  
 GN YPFLC1.  
 OS Chlamydomonas reinhardtii.  
 CC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 CC Chlamydomonadaceae; Chlamydomonas.  
 CX NCB1\_TaxID-3055;  
 RX NCBI\_TaxID-3055;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CW15;  
 RX MEDLINE-95309723; PubMed-7789809;  
 RA Dietmaier W., Fabry S., Huber H., Schmitt R.;

RT "Analysis of a family of ypt genes and their products from  
 RT Chlamydomonas reinhardtii.";  
 RL Gene 158:41-50(1995).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U13168; AAB2727.1;  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsmfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF000071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRMNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfam: TIGR00231; small\_GTP; 1.  
 DR GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 KW Multigene family.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 203 203 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 203 AA; 22599 MW; DD22420FE2F35E1D CRC64;  
 Query Match 73.0%; Score 646.5; DB 1; Length 203;  
 Best Local Similarity 65.7%; Pred. No. 5.3e-47;  
 Matches 134; Conservative 9; Mismatches 26; Indels 35; Gaps 3;  
 QY 4 MNPEYDYELFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDGKTIKQ 63  
 DB 1 MNPEYDYELFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDFKRTIELDGKTIKQ 60  
 QY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91  
 DB 61 IMPYAGQERFRTITSSYRGAHGICVYDVTDMDSFNNVKOMLOEIDRYATEGVNKLIVG 120  
 QY 92 NKCDLTKRKVVDYTTAKAFADSLGIPLETSAKNATNVEGSPFMMAAEIKRKGSCATAG 149  
 DB 121 NKSDMTKAKVVEYTAKEFADSLGIPLETSAKNATNVEGAFETMARQIKEMSGSTAIN 180  
 QY 150 AGAERKSNVKIQTSPVKOSGGCC 173  
 DB 181 KAGGPVVRPO-BGRPINSKSSCC 203  
 RESULT 8  
 YPFL\_SCHPO STANDARD; PRT; 203 AA.  
 ID YPFL\_SCHPO  
 AC P11620;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ras-related protein YPFL.  
 GN YPFL OR SPBC1703.10.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetaceae; Schizosaccharomycetaceae;  
 CC Schizosaccharomyces.  
 CX NCB1\_TaxID-4896;  
 RX NCBI\_TaxID-4896;  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-90228338; PubMed-2328721;  
 RA Miyake S., Yamamoto M.;

RT \*Identification of ras-related, YPT family genes in  
 RT Schizosaccharomyces pombe.\*;  
 RL EMBL J. 9:1417-1422(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-89296485; PubMed-2740223;  
 RA Fawell E., Hook S., Armstrong J.;  
 RT \*Nucleotide sequence of a gene encoding a YPT1-related protein from  
 RT Schizosaccharomyces pombe.\*;  
 RL Nucleic Acids Res. 17:4373-4373(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RX MEDLINE-21848401; PubMed-11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouris J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares K., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,  
 RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Mambutti R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Delaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garcon A., Rhode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;  
 RT \*The genome sequence of Schizosaccharomyces pombe.\*;  
 RL Nature 415:871-880(2002).  
 RN [4]  
 RP ISOPRENOL.  
 RX MEDLINE-92283842; PubMed-1597466;  
 RA Newman C.M., Giannakouris T., Hancock J.F., Fawell E.H., Armstrong J.,  
 RA Magee A.I.;  
 RT \*Post-translational processing of Schizosaccharomyces pombe YPT  
 RT proteins.\*;  
 RL J. Biol. Chem. 267:11329-11336(1992).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X52099; CA36319.1; -;  
 DR EMBL: X15082; CA33192.1; -;  
 DR EMBL: AL136536; CAB66454.1; -;  
 DR PIR: S04590; S04590.  
 DR PIR: S10025; S10025.  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfams: TIGR00231; small\_GTP; 1.

KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 202 202 GERANYL-GERANYL.  
 FT LIPID 203 203 GERANYL-GERANYL.  
 FT CONFLICT 1 4 MNPE -> MNANINR (IN REF. 2).  
 SQ SEQUENCE 203 AA; 22816 MW; 59B3524F42B3094 CRC64;  
 Query Match 73.0%; Score 646.5; DB 1; Length 203;  
 Best Local Similarity 64.5%; Pred. No. 5.3e-47;  
 Matches 131; Conservative 16; Mismatches 23; Indels 33; Gaps 2;  
 QY 4 MNPEVDYLFKLLIDSGGKSCLLRPADPDYTESYSTIGVDPRKIRIETIDGKTIKQ 63  
 DB 1 MNPEVDYLFKLLIDSGGKSCLLRPADPDYTESYSTIGVDPRKIRIETIDGKTIKQ 60  
 QY 64 I-----ESFNKYKWLQIDRYASENVKLLVG 91  
 DB 61 IWDTAGQERFRTTSSYYRGAGHIIIVYDVTQDSFNNVKWLQIDRYAVGVNRLVG 120  
 QY 92 NKCDLTKKYVDYTAKEPADSLGIPLETSAKNTNVDQSMYMAAETKRMGCATAG 151  
 DB 121 NKSDVDKKVVEYSYAKEPADSLNIPLETSAKSDSTNVDQAFLTMSROIKEKMGNTTAS 180  
 QY 152 GAERSNVXI-OSTPVKSGGCCC 173  
 DB 161 SNAKSVAVGQGTNNVSQSSNCC 203  
 RESULT 9  
 ID YPT1\_VOLCA STANDARD; PRT; 203 AA.  
 AC P31584;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GTP-binding protein ypt1.  
 GN YPTV1.  
 OS Volvox carterii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Volvocaceae; Volvox.  
 OX NCBI\_TaxID=3067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-f. Nagariensis / HK10;  
 RX MEDLINE-92380499; PubMed-1511889;  
 RA Fabry S., Nass N., Huber H., Palme K., Jaenicke L., Schmitt R.;  
 RT \*The ypt1 gene encodes a small G-protein in the green alga Volvox  
 RT carterii: gene structure and properties of the gene product.\*;  
 RL Gene 118:153-162(1992).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M93438; AAA34255.1; -;  
 DR PIR: JC1247; JC1247.  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.

DR TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 MW Multi-gene family.  
 FT NE\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NE\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NE\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 203 203 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 203 AA; 22532 MW; CD143BADB92B4C65 CRC64;

Query Match 72.5%; Score 642.5; DB 1; Length 203;  
 Best Local Similarity 65.0%; Pred. No. 1,1e-46;  
 Matches 134; Conservative 6; Mismatches 27; Indels 39; Gaps 3;

DR 4 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESTYSTIGVDKRTITELDKRTIKQ 63  
 DB 1 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESTYSTIGVDKRTITELDKRTIKQ 60  
 QY 64 I-----ESFNNVKOMLOEIDRYASENVNKLVG 91  
 DB 61 IWDNAGEFRRTITSSYRGAAGIIVYDVTDOESFNNVKOMLAIEDRYASENVNKLVG 120  
 QY 92 NKCDLTKKVVDTYTAKEFADSLGIPLETSAKKNATNEQSFMTAAEIKRKGPGATNG 151  
 DB 121 NKSDLTAKKVVSTDAKEFAESLGIEPLETSKKNATNEQAFMTAAEIKRKGPGATNG 180  
 QY 148 ATAGGAESKNVXIQSTPVKQSGGACC 173  
 DB 181 KPGGPVVRPT--EGKPTNNKSSCC 203

RESULT 10  
 YPT1\_PHYIN STANDARD; PRT; 201 AA.  
 AC Q01890;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Ras-like GTP-binding protein YPT1.  
 GN Phytophthora infestans (Potato late blight fungus).  
 OS Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
 OC Phytophthora.  
 NC NCBI\_TaxID=4787;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 52009;  
 RX MEDLINE=97128773; PubMed=8973313;  
 RA Chen Y, Rozdy R.  
 RT "Characterization of a Phytophthora infestans gene involved in vesicle transport";  
 RL Gene 181:89-94(1996).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR TRAFFIC FROM ER TO GOLGI (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC EMBL: U30474; AAB40355.1; -  
 CC HSSP: P05713; 3BAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_transfmg.  
 DR Pfam: PFO0071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRANG.  
 SMART: SM00175; RAB; 1.

DR TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 MW Multi-gene family.  
 FT NE\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NE\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NE\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 200 200 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 201 AA; 22292 MW; 89EE53A693C8AC5 CRC64;

Query Match 71.3%; Score 631.5; DB 1; Length 201;  
 Best Local Similarity 67.5%; Pred. No. 9.2e-46;  
 Matches 139; Conservative 8; Mismatches 18; Indels 41; Gaps 5;

DR 4 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESTYSTIGVDKRTITELDKRTIKQ 63  
 DB 1 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESTYSTIGVDKRTITELDKRTIKQ 60  
 QY 64 I-----ESFNNVKOMLOEIDRYASENVNKLVG 91  
 DB 61 IWDNAGEFRRTITSSYRGAAGIIVYDVTDOESFNNVKOMLAIEDRYASENVNKLVG 120  
 QY 92 NKCDLTKKVVDTYTAKEFADSLGIPLETSAKKNATNEQSFMTAAEIKRKGPGATNG 151  
 DB 121 NKSDLTAKKVVSTDAKEFAESLGIEPLETSKKNATNEQAFMTAAEIKRKGPGATNG 177  
 QY 152 GAESKNVXIQSTPVKQ--SGG--CC 173  
 DB 178 VAFRAGVKL--TPGQVPSNGSGKCC 201

RESULT 11  
 ARAS\_ARATH STANDARD; PRT; 258 AA.  
 AC P28188;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 15-JUN-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ras-related protein ARAS-5.  
 GN Arabidopsis thaliana (Mouse-ear cress).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; eudicotyledons; Core eudicots; Rosidae;  
 NC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altieri J., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hulzar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marshall A., Millschott J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Utechtbeck T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana".  
 RL Nature 408:816-820(2000).  
 RN [2]  
 RP SEQUENCE OF 64-258 FROM N.A.  
 RC STRAIN=Various strains; TISSUE=leaf;  
 RX MEDLINE=92084144; PubMed=1748311;  
 RA Anal T., Hasegawa K., Watanabe Y., Uchimiya H., Ishizaki R.,

RA Matsui M.;  
 RT "Isolation and analysis of cDNAs encoding small GTP-binding proteins  
 of Arabidopsis thaliana.";  
 RL Gene 108:259-264(1991).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U89959; AAC24370.1; -  
 CC EMBL: D01027; BAA00832.1; -  
 CC PIR: PS0279; PS0279.  
 CC HSSP: P05713; 3RAB.  
 CC InterPro: IPR003579; GTPase\_Rab.  
 CC InterPro: IPR001806; Ras\_trnsfmrng.  
 CC Pfam: PF00071; ras; 1.  
 CC PRINTS: PRO0449; RASTRNSFRNG.  
 CC SMART: SM00175; RAB; 1.  
 CC TIGRFAMs: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Multigene family.  
 FT NP\_BIND 70 77 GTP (BY SIMILARITY).  
 FT NP\_BIND 118 122 GTP (BY SIMILARITY).  
 FT NP\_BIND 176 179 GTP (BY SIMILARITY).  
 FT DOMAIN 92 100 EFFECTOR REGION (BY SIMILARITY).  
 FT LIPID 255 255 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 256 256 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 258 AA; 29169 MW; 34D278A0C504638F CRC64;  
 Query Match 69.9%; Score 619.5; DB 1; Length 258;  
 Best Local Similarity 61.6%; Pred. No. 1.2e-44;  
 Matches 125; Conservative 15; Mismatches 30; Indels 33; Gaps 2;  
 QY 3 MNPEYDYLKLLIGDSGVGSKLLRFADDTYESTYSTIGVDFKIRTELDGKTIKL 62  
 DB 55 TMNPEYDYLKLLIGDSGVGSKLLRFSDSYESTYSTIGVDFKIRTELDGKTIKL 114  
 QY 63 QI-----ESFNNVQWLOEIDRYASENVNKLIV 90  
 DB 115 QIMDTAGOEFRRTTSSYYRGAGIILYDVTDSEFNNVQWLOEIDRYASDNVKNLIV 174  
 QY 91 GNKCDLTKKVVDTTAKFADSLGIPLETSAKNATVQESFMTMAEIKRMGPAGATA 150  
 DB 175 GNKCDLTKKVVDTTAKFADSLGIPLETSAKNATVQESFMTMAEIKRMGPAGATA 150  
 QY 151 GGAESKNVQISTPVKQSGGCC 173  
 DB 235 NNARPTVQIRGQPVNOK-NGCC 256  
 RESULT 12  
 ID YPT2\_MAIZE STANDARD; PRT; 203 AA.  
 AC Q05737;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GTP-binding protein YPT2.  
 GN YPT2.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Palicoudeae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RX SEQUENCE FROM N.A.  
 RP TISSUE=Coleoptile;  
 RZ MEDLINE=92115746; PubMed=1731354;

RA Palme K., Diefenthal T., Vingron M., Sander C., Schell J.;  
 RT "Molecular cloning and structural analysis of genes from Zea mays  
 (L.) coding for members of the ras-related ypt gene family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:787-791(1992).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: ITS EXPRESSION IS WEAK IN STEMS, HIGHER IN  
 CC ROOTS, LEAVES AND COLEOPTILES, BUT HIGHEST IN FLOWERS.  
 CC -1- PFM: THE PALMITIC ACID IS REQUIRED FOR MEMBRANE ATTACHMENT AND  
 CC BIOLOGICAL FUNCTION.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X63278; CAA44919.1; -  
 CC PIR: B38202; B38202.  
 CC HSSP: P05713; 3RAB.  
 CC MaizeDB: 78605; -  
 CC InterPro: IPR003579; GTPase\_Rab.  
 CC InterPro: IPR001806; Ras\_trnsfmrng.  
 CC Pfam: PF00071; ras; 1.  
 CC PRINTS: PRO0449; RASTRNSFRNG.  
 CC SMART: SM00175; RAB; 1.  
 CC TIGRFAMs: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Palmitate; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 200 200 PALMITATE (BY SIMILARITY).  
 FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 203 AA; 22475 MW; E2413267ACD1B8A CRC64;  
 Query Match 69.8%; Score 618.5; DB 1; Length 203;  
 Best Local Similarity 63.4%; Pred. No. 1.1e-44;  
 Matches 128; Conservative 11; Mismatches 30; Indels 33; Gaps 2;  
 QY 4 MNPEYDYLKLLIGDSGVGSKLLRFADDTYESTYSTIGVDFKIRTELDGKTIKL 63  
 DB 1 MNPEYDYLKLLIGDSGVGSKLLRFADDTYESTYSTIGVDFKIRTELDGKTIKL 60  
 QY 64 I-----ESFNNVQWLOEIDRYASENVNKLIV 91  
 DB 61 IMDTAGOEFRRTTSSYYRGAGIILYDVTDSEFNNVQWLOEIDRYASDNVKNLIV 120  
 QY 92 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATVQESFMTMAEIKRMGPAGATA 151  
 DB 121 NKSDLTANKVATETAKAFADSLGIPLETSAKNATVQAFMAASIDRMASQPAAL 180  
 QY 152 GAESKNVQISTPVKQSGGCC 173  
 DB 181 NARAPTQIRGQPVNOK-TSCC 201  
 RESULT 13  
 ID RIC1\_ORYSA STANDARD; PRT; 202 AA.  
 AC P40392;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ras-related protein RIC1.  
 GN RIC1.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;



```

FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).
FT LIPID 205 205 PALMITATE.
FT LIPID 206 206 GERANYL-GERANYL.
FT MUTAGEN 37 37 Y->F: NO CHANGE.
FT MUTAGEN 39 39 S->A: NO CHANGE.
FT MUTAGEN 40 40 T->S: NO CHANGE.
FT MUTAGEN 41 41 I->M: LETHAL.
FT MUTAGEN 43 43 V->E: NO CHANGE.
FT MUTAGEN 44 44 D->N: TEMPERATURE-SENSITIVE PHENOTYPE.
FT MUTAGEN 121 121 N->I: LETHAL.
FT CONFIDENT 171 171 E->Q (IN REF. 1).
SQ SEQUENCE 206 AA: 23214 MW: F8C704F6BF2D227B CRC64;

Query Match 64.4%; Score 571; DB 1; Length 206;
Best Local Similarity 56.3%; Pred. No. 1e-40;
Matches 116; Conservative 21; Mismatches 33; Indels 36; Gaps 2;

QY 4 MNPEYDYLKLLIDSGVSKSCLLRADPTYSYSTIGVDFKIRTIELDKTKIKQ 63
DB 1 MNPEYDYLKLLIDSGVSKSCLLRADPTYSYSTIGVDFKIRTIELDKTKIKQ 60
QY 64 I-----ESFNNVQWLOEIDRYASENVKLLVG 91
DB 61 IWDTAGQERFRITSSYYRGAGHIIYVDYDKLFEFNVQWLOEIDRYASENVKLLVG 120
QY 92 NKCDLITKKVVDYTTAKFADSLGIPLETSAKNTNVEOSFMTAAEIKRMGP-----G 147
DB 121 NKCDLITKKVVDYTTAKFADSLGIPLETSAKNTNVEOSFMTAAEIKRMGP-----G 180
QY 148 ATGAGAEKSNVKIQSTPVKQSGGCC 173
DB 181 TTQKKEDKGNVNLKGOSLNTGCGCC 206

RESULT 15
RBLA_DICDI STANDARD; PRT; 167 AA.
AC P34139;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Ras-related protein RAB1A.
GN RAB1A OR RAB1A.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycelozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=94124041; PubMed=8294041;
RA Bush J.M. IV, Franek K., Daniel J.M., Spiegelman G.B., Weeks G.,
RT Cardelli J.A.;
RT "Cloning and characterization of five novel Dictyostelium discoideum
RT rad-related genes.";
RL Gene 136:55-60(1993).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L21009; AAC37385.1; -.
CC HSSP; P05713; 3RAB.
CC DictyDb; DD05055; raba.
CC InterPro; IPR003579; GTPase_Rab.
CC InterPro; IPR001806; Ras_trnsfmg.

```

```

DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; Ras; 1.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
KW GTP-binding.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
SQ SEQUENCE 167 AA: 18943 MW: 72480B04D943BD8D CRC64;

Query Match 61.2%; Score 542; DB 1; Length 167;
Best Local Similarity 67.7%; Pred. No. 2e-38;
Matches 111; Conservative 10; Mismatches 11; Indels 32; Gaps 1;

QY 4 MNPEYDYLKLLIDSGVSKSCLLRADPTYSYSTIGVDFKIRTIELDKTKIKQ 63
DB 1 MNPEYDYLKLLIDSGVSKSCLLRADPTYSYSTIGVDFKIRTIELDKTKIKQ 60
QY 64 I-----ESFNNVQWLOEIDRYASENVKLLVG 91
DB 61 IWDTAGQERFRITSSYYRGAGHIIYVDYDKLFEFNVQWLOEIDRYASENVKLLVG 120
QY 92 NKCDLITKKVVDYTTAKFADSLGIPLETSAKNTNVEOSFMT 135
DB 121 NKSDLVARKVVDENTAKAFADSLQIPLETSAKNTNVEOAFMT 164

Search completed: June 18, 2003, 15:21:38
Job time : 23 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 15:18:29 ; Search time 80 Seconds

(without alignments)  
445,577 Million cell updates/sec

Title: US-09-820-003b-2

Perfect score: 1 MSSMNPEDYLFKLLIGDS.....EKSNNKIQSTPYKSGGCGC 173

Sequence: Sequence: 1 MSSMNPEDYLFKLLIGDS.....EKSNNKIQSTPYKSGGCGC 173

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- 1: SP\_ARCHAEA:\*
- 2: SP\_BACTERIA:\*
- 3: SP\_FUNGI:\*
- 4: SP\_HUMAN:\*
- 5: SP\_INVERTEBRATE:\*
- 6: SP\_MAMMAL:\*
- 7: SP\_MHC:\*
- 8: SP\_ORGANELLE:\*
- 9: SP\_PHAGE:\*
- 10: SP\_PLANT:\*
- 11: SP\_PODENT:\*
- 12: SP\_VIRUS:\*
- 13: SP\_VERTEBRATE:\*
- 14: SP\_UNCLASSIFIED:\*
- 15: SP\_VIRUS:\*
- 16: SP\_BACTERIAP:\*
- 17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	766.5	86.5	201	4 Q9H0U4	Q9H0U4 homo sapien
2	759.5	85.7	201	11 Q9D1G1	Q9D1G1 mus musculu
3	700	79.0	141	4 Q9BNE1	Q9BNE1 homo sapien
4	683	77.1	202	5 Q183J2	Q183J2 bombyx mori
5	671	75.1	202	5 Q161A3	Q161A3 bombyx mori
6	660	74.5	203	3 Q9UQA6	Q9UQA6 caenorhabdi
7	653.5	74.0	201	3 Q9HEH3	Q9HEH3 aspergillus
8	653.5	73.8	201	3 Q9P8J7	Q9P8J7 aspergillus
9	644	72.5	202	10 Q40Z03	Q40Z03 lotus japon
10	640	72.5	202	10 Q9SEH3	Q9SEH3 arabidopsis
11	636.5	71.8	202	3 Q9H0T5	Q9H0T5 trichoderma
12	634.5	71.6	202	10 Q8M1S4	Q8M1S4 pisum sativ
13	633	71.4	202	10 Q8M4S8	Q8M4S8 arabidopsis
14	632.5	71.4	202	10 Q9FPJ4	Q9FPJ4 arabidopsis
15	632.5	71.3	203	10 Q24112	Q24112 nicotiana p
16	632	71.3	202	10 Q9SXT5	Q9SXT5 cicor arlet

17	628.5	70.9	203	10 Q40Z04	Q40Z04 lotus japon
18	627.5	70.8	203	10 Q40569	Q40569 nicotiana t
19	624.5	70.5	203	10 Q40205	Q40205 lotus japon
20	624	70.4	129	4 Q9Y3T2	Q9Y3T2 homo sapien
21	623	70.3	202	10 Q39845	Q39845 glycine max
22	622.5	70.3	203	10 Q9M7P5	Q9M7P5 capsicum an
23	622.5	70.3	203	10 Q41338	Q41338 lycopersico
24	622.5	70.3	203	10 Q41340	Q41340 lycopersico
25	621.5	70.1	203	10 Q949E2	Q949E2 oryza sativ
26	619	69.9	202	10 Q08155	Q08155 pisum sativ
27	618.5	69.8	202	10 Q940Z7	Q940Z7 arabidopsis
28	616	69.5	202	10 Q08153	Q08153 pisum sativ
29	612	69.1	221	10 Q23594	Q23594 arabidopsis
30	608	68.6	203	10 Q8RU63	Q8RU63 oryza sativ
31	598	67.5	218	10 Q39861	Q39861 glycine max
32	596.5	67.3	207	3 Q96W97	Q96W97 candida alb
33	595.5	66.6	202	10 Q9ZRH6	Q9ZRH6 petunia hyb
34	589	66.5	201	10 Q08152	Q08152 pisum sativ
35	589	66.5	451	10 Q9LGG3	Q9LGG3 oryza sativ
36	587	66.3	201	10 Q43462	Q43462 glycine max
37	583.5	65.9	202	10 Q41339	Q41339 lycopersico
38	581	65.6	201	10 Q40201	Q40201 lotus japon
39	580	65.5	200	5 Q253Z4	Q253Z4 leishmania
40	574	64.8	205	10 Q38Z74	Q38Z74 brassica ca
41	569.5	64.3	205	10 Q3ZKE2	Q3ZKE2 arabidopsis
42	565	64.2	194	10 Q40202	Q40202 lotus japon
43	563	63.3	206	3 Q9HRM5	Q9HRM5 kluyveromyc
44	503	56.8	200	5 Q9BHN1	Q9BHN1 plasmodium
45	484	54.6	205	5 Q9BLF3	Q9BLF3 entamoeba h

ALIGNMENTS

RESULT 1	ID	Q9H0U4	PRELIMINARY:	PRT:	201 AA.
AC	Q9H0U4	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Hypothetical 22.2 kDa protein.				
GN	DKFP5641172.				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BRAIN.				
RX	MEDLINE-2154917; PubMed-11230166;				
RA	Wiemann S., Well B., Wellenreuther R., Gassenhuber J., Glasl S.,				
RA	Ansgore W., Boscher M., Bloeker H., Baerends S., Blum H.,				
RA	Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,				
RA	Mewes H.W., Oltmannseder B., Obermaier B., Tampe J., Heuner D.,				
RA	Wamulit R., Korn B., Klein W., Poustka A.,				
RT	Towards a Catalog of Human Genes and Proteins: Sequencing and				
RT	Analysis of 500 Novel Complete Protein Coding Human cDNAs.*;				
RL	Genome Res. 11:422-435(2001).				
CC	-1- SIMILARITY: TO RAB PROTEINS. BELONGS TO THE RAB SUBFAMILY.				
EMBL	AL136635; CAB6570.1; -				
DR	HSSP; P05713; JRA8.				
DR	InterPro; IPR003579; GTPase_Rab.				
DR	InterPro; IPR001806; Ras_Transfmg.				
DR	InterPro; IPR002078; Sig54_Interact.				
DR	InterPro; IPR005225; Small_GTP.				
DR	Pfam; PF00071; Ras; 1.				
DR	PRINTS; PR00449; RASTRNSFRNG.				
DR	SMART; SM00175; RAB; 1.				
DR	TIGRFAMS; TIGR00231; small_GTP; 1.				
DR	PROSITE; PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.				
KW	GTP-binding; Hypothetical protein; Lipoprotein.				
SO	SEQUENCE 201 AA; 22171 MW; 9812FF4DAC34B2BE CRC64;				

Query Match 86.5%; Score 766.5; DB 4; Length 201;  
 Best Local Similarity 77.2%; Pred. No. 4,1e-54;  
 Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

QY 4 MNPEYDYLFKLLIGDSGVGKSCLLRFADPTYESYSTIGVDKFIETILDGKTIQ 63  
 DB 1 MNPEYDYLFKLLIGDSGVGKSCLLRFADPTYESYSTIGVDKFIETILDGKTIQ 60  
 QY 64 I-----ESFNNVKQWLOEIDRYASENNKLLVG 91  
 DB 61 IWDTAGGERFTITSSYRGAGHIIIVYDVTDQESYANVKQWLOEIDRYASENNKLLVG 120  
 QY 92 NKCDLTKKVVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTAAEIKRMGPGATAG 151  
 DB 121 NKSDLTTRKKVDDNTAKFADSLGIPLETSAKNATNVEQAFMTAAEIKRMGPGASG 180  
 QY 152 GAESNVKIOSTPVKQSGGCC 173  
 DB 181 G-ERPNLKIDSTPVKPGAGGCC 201

## RESULT 2

Q9DIG1

ID Q9DIG1 PRELIMINARY; PRT; 201 AA.

AC Q9DIG1 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE 1110011F09RIK protein (RIKEN CDNA 1110011F09 gene).  
 GN 1110011F09RIK  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRIO;  
 RX MEDLINE=21085600; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,  
 RA Saito T., Okazaki Y., Gojohori T., Hono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Pletschmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,  
 RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guenlinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Sasaki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N. A.  
 RC TISSUE=KIDNEY;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.  
 DR EMBL: AK003609; BAB22888.1; -;  
 DR EMBL: BC016408; AAH16408.1; -;  
 DR HSSP: P05713; 3RAB.  
 DR MGD: MGI:1923558; 1110011F09RIK.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR InterPro: IPR002078; Sig54\_interact.  
 DR InterPro: IPR005225; Small\_GTP.

DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PRO0449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAMS: TIGR00231; small\_gtp; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding; lipoprotein.  
 SQ SEQUENCE 201 AA; 22187 MW; 870DF52AEF4B2BE CRC64;

Query Match 85.7%; Score 759.5; DB 11; Length 201;  
 Best Local Similarity 76.2%; Pred. No. 1,5e-53;  
 Matches 154; Conservative 7; Mismatches 8; Indels 33; Gaps 2;

QY 4 MNPEYDYLFKLLIGDSGVGKSCLLRFADPTYESYSTIGVDKFIETILDGKTIQ 63  
 DB 1 MNPEYDYLFKLLIGDSGVGKSCLLRFADPTYESYSTIGVDKFIETILDGKTIQ 60  
 QY 64 I-----ESFNNVKQWLOEIDRYASENNKLLVG 91  
 DB 61 IWDTAGGERFTITSSYRGAGHIIIVYDVTDQESYANVKQWLOEIDRYASENNKLLVG 120  
 QY 92 NKCDLTKKVVDDYTTAKFADSLGIPLETSAKNATNVEQSFMTAAEIKRMGPGATAG 151  
 DB 121 NKSDLTTRKKVDDNTAKFADSLGIPLETSAKNATNVEQAFMTAAEIKRMGPGASG 180  
 QY 152 GAESNVKIOSTPVKQSGGCC 173  
 DB 181 G-ERPNLKIDSTPVKPGAGGCC 201

## RESULT 3

Q96N61

ID Q96N61 PRELIMINARY; PRT; 141 AA.

AC Q96N61 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE CDNA FLJ33365 fis, clone NB9M4100135, highly similar to RAS-related  
 DE protein RAB-1A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RA Taashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,  
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,  
 RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K.,  
 RA Isoqal T.;  
 RT "NEO human cDNA sequencing project.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK055927; BAB71048.1; -;  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR InterPro: IPR002078; Sig54\_interact.  
 DR Pfam: PF00071; ras; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding.  
 SQ SEQUENCE 141 AA; 15330 MW; 620F586CA34BA77 CRC64;

Query Match 79.0%; Score 700; DB 4; Length 141;  
 Best Local Similarity 81.5%; Pred. No. 5,8e-49;  
 Matches 141; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MSSNPEYDYLFKLLIGDSGVGKSCLLRFADPTYESYSTIGVDKFIETILDGKTI 60  
 DB 1 MSSNPEYDYLFKLLIGDSGVGKSCLLRFADPTYESYSTIGVDKFIETILDGKTI 32  
 QY 61 KLOIESFNNVKQWLOEIDRYASENNKLLVGKCDLTKKVVDDYTTAKFADSLGIPLET 120  
 DB 33 -----ESFNNVKQWLOEIDRYASENNKLLVGKCDLTKKVVDDYTTAKFADSLGIPLET 88



```

QY      121 TSANKATNVQSPMTMAEIKKRGKPGATGAGKSNKTIQSTPYKOSGGGCC 173
DB      89 TSANKATNVQSPMTMAEIKKRGKPGATGAGKSNKTIQSTPYKOSGGGCC 141

RESULT 4
ID      018332      PRELIMINARY;      PRT;      205 AA.
AC      018332;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      RAB1 protein.
GN      RAB1 OR C63320.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BERKELEY;
RX      MEDLINE=20196006, PubMed=10731132;
RA      Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Wortman J.R., Fendell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers J.H.C., Blake J.R., Champagne M., Pfeiffer B.D.,
RA      An H.-J., Andrews-Plannkoch C., Balwin D.,
RA      Arill J.F., Ashby A., An H.-J., Andrews-Plannkoch C., Balwin D.,
RA      Balw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beaslev E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA      Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Fosler C., Gebrelian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA      Gload A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
RA      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Mikhlov G., Milehna N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA      Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA      Slater E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Syrakas R., Tector C., Turner R., Venier E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisenbach J.,
RA      Williams S.M., Woodage T., Wooley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yen R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA      Gbbs R.A., Myers E.W., Rubin M., Ventre J.C.;
RT      *The genome sequence of Drosophila melanogaster.*;
RL      Science 287:2185-2195(2000).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=OREGON R; TISSUE=HEAD;
RX      MEDLINE=97228579, PubMed=9074639;
RA      Satoh A.K., Tokunaga F., Ozaki K.;
RT      *Rab proteins of Drosophila melanogaster: novel members of the Rab-
RT      protein family.*;
RL      FEBS Lett. 404:65-69(1997).
CC      -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR      EMBL: AF003734; AAF55873.1;
DR      EMBL: D84312; BAA21705.1;

```

---

```

DR      HSP: P05713; 3RAB.
DR      Flybase: FBgn016700; Rab1.
DR      InterPro: IPR003579; GTPase_Rab.
DR      InterPro: IPR001806; Ras_trnsfmg.
DR      InterPro: IPR002078; S1954_interact.
DR      InterPro: IPR005225; Small_GTP.
DR      Pfam: PF00071; Ras; 1.
DR      PRINTS: PRO0449; RASTRNSFRNG.
DR      SMART: SM00175; RAB; 1.
DR      TIGRFAMs: TIGR00231; small_gtp; 1.
DR      PROSITE: PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
DR      GTP-binding; Lipoprotein.
SQ      SEQUENCE 205 AA; 22762 MW; 9CE2691205FAE343 CRC64;

Query Match      77.1%; Score 683; DB 5; Length 205;
Best local similarity 69.4%; Pred. No. 2, 2e-47;
Matches 143; Conservative 12; Mismatches 17; Indels 34; Gaps 3;

QY      1 MSSMNPVYTLKLLIGDSGVKSCLLLRPADTTESTISTIGYDKRTITLDGKTI 60
DB      1 MSSVNPDIYTLKLLIGDSGVKSCLLLRPADTTESTISTIGYDKRTITLDGKTI 60
QY      61 KLOI-----ESFNNVKQLQELIDRYASENVNKL 88
DB      61 KLOIWPAGQEFRTITSSYRGAGIIIVYDCIDQESFNNVKQWLEIEERYACENVNKL 120
QY      89 LVGNKCDLTKRVVDTTAKFADSLGIPLETSAKNATNVQSPMTMAEIKKRGKPGQA 148
DB      121 LVGNKCDLTKRVVDTTAAVYAQLGIPLETSAKSATNVQSPMTMAEIKKRGKPGPS 180
QY      149 TAGAGKSNVKI-QSTPYKOSGGGCC 173
DB      181 SARD-NASRVKIDQGRVEMTKSGCC 205

RESULT 5
ID      016143      PRELIMINARY;      PRT;      202 AA.
AC      016143;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Small GTP-binding protein.
GN      RAB8.
OS      Bombyx mori (Silk moth).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC      Bombycoidea; Bombycidae; Bombyx.
OX      NCBI_TaxID=7091;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      DNO T.; Nakajima A., Ueno M., Aizono Y.;
RT      *Isolation of cDNA encoding small GTP-binding protein from Bombyx
RT      mori.*;
RL      Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
DR      EMBL: AF013572; AAB67169.1;
DR      HSP: P05713; 3RAB.
DR      InterPro: IPR003579; GTPase_Rab.
DR      InterPro: IPR001806; Ras_trnsfmg.
DR      InterPro: IPR002078; S1954_interact.
DR      InterPro: IPR005225; Small_GTP.
DR      Pfam: PF00071; Ras; 1.
DR      PRINTS: PRO0449; RASTRNSFRNG.
DR      SMART: SM00175; RAB; 1.
DR      TIGRFAMs: TIGR00231; small_gtp; 1.
DR      PROSITE: PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
DR      GTP-binding; Lipoprotein.
SQ      SEQUENCE 202 AA; 22364 MW; 0338E8950630DFP4 CRC64;

Query Match      75.7%; Score 671; DB 5; Length 202;
Best local similarity 66.5%; Pred. No. 2e-46;
Matches 135; Conservative 16; Mismatches 16; Indels 34; Gaps 3;

```

```

OY 4 NNPEDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGTIKLQ 63
DB 1 NNPEDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGTIKLQ 60
OY 64 I-----ESFNNYKWLQEIIDRYASENNKLLVG 91
DB 61 IMDTAGOERFRTITSSYRGAGIIIVYDCTQDSEFSNNKWLQEIIDRYACNNKLLVG 120
OY 92 NNCDDLTTRKVVDTYTAKEPADSLGIPLETSAKNATNVQSFMTAAETKRMGPGATAG 151
DB 121 NNCDDLTTRKVVDTYTAKEPADSLGIPLETSAKNATNVQSFMTAAETKRMGPGATAG 179
OY 152 GAESKNVXI-QSTPVKQSGGGCC 173
DB 180 AAPAGHVKIDGGQPIDTGKSSCC 202

RESULT 6
OY 09UA06 PRELIMINARY; PRT; 205 AA.
AC 09UA06;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DE 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 22.5 kDa protein.
GN C39F7.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RT Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Scheet P., Magg L., Dubbelde C.;
RT "The sequence of C. elegans cosmid C39F7."
RT Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RT Submitted (Sep-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.
DB EMBL; AF101310; AAC69218.1; -.
DR HSSP; P01112; 1PLI.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmg.
DR InterPro; IPR002078; S1954_interact.
DR InterPro; IPR005225; Sma11_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMS; TIGR00231; sma11_GTP.1.
DR PROSITE; PS00675; SIGMAS4_INTERACT_1; UNKNOWN_1.
KW GTP-binding; Hypothetical protein; Lipoprotein.
SQ SEQUENCE 205 AA; 22545 MW; 5221920FD05842D3 CRC64;

Query Match 74.5%; Score 660; DB 5; Length 205;
Best Local Similarity 66.2%; Pred. No. 1.5e-45;
Matches 137; Conservative 16; Mismatches 18; Indels 36; Gaps 4;

OY 1 MSSNPEDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGTIKI 60
DB 1 MAANNPEDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGTIKI 60

```

```

OY 61 KIQI-----ESFNNYKWLQEIIDRYASENNKLLVG 88
DB 61 KIQIIMDTAGOERFRTITSSYRGAGIIIVYDCTQDSEFSNNKWLQEIIDRYACNNKLLVG 120
OY 89 LVNKKDDLTTRKVVDTYTAKEPADSLGIPLETSAKNATNVQSFMTAAETKRMGPGCA 148
DB 121 LVNKKDDLTTRKVVDTYTAKEPADSLGIPLETSAKNATNVQSFMTAAETKRMGPGCA 180
OY 149 TAGGAESKNVXI-QSTPVKQSGGGCC 173
DB 181 GAGGA--PGVITGSGQPVQDKKSGGCC 205

RESULT 7
OY 09HET3 PRELIMINARY; PRT; 201 AA.
AC 09HET3;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DE 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Secretion related GTPase (SrgB).
GN SRGB.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_Taxid=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-N402;
RA Punt P.J., Selbooth B., Weenl X.O., van Zel J.C.M., Lenders M.,
RA Konetschny C., Ram A.F., Montijn R., Kubicek C.P.,
RA Van Den Hondel C.A.;
RT "Identification and characterisation of a family of secretion related
RT small GTPase encoding genes from the filamentous fungus Aspergillus
RT niger: a putative SEC4 homologue is not essential for growth."
RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.
DB EMBL; AJ278659; CAC17833.1; -.
DR HSSP; P01112; 1PLI.
DR InterPro; IPR003579; GTPase_Rab.
DR InterPro; IPR001806; Ras_trnsmg.
DR InterPro; IPR002078; S1954_interact.
DR InterPro; IPR005225; Sma11_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00175; RAB; 1.
DR TIGRFAMS; TIGR00231; sma11_GTP.1.
DR PROSITE; PS00675; SIGMAS4_INTERACT_1; UNKNOWN_1.
KW GTP-binding; Lipoprotein.
SQ SEQUENCE 201 AA; 22303 MW; 8DC1A5B0B8BDF378 CRC64;

Query Match 74.0%; Score 655.5; DB 3; Length 201;
Best Local Similarity 67.2%; Pred. No. 3.4e-45;
Matches 137; Conservative 12; Mismatches 18; Indels 37; Gaps 3;

OY 4 NNPEDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGTIKLQ 63
DB 1 NNPEDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGTIKLQ 60
OY 64 I-----ESFNNYKWLQEIIDRYASENNKLLVG 91
DB 61 IMDTAGOERFRTITSSYRGAGIIIVYDCTQDSEFSNNKWLQEIIDRYACNNKLLVG 120
OY 92 NNCDDLTTRKVVDTYTAKEPADSLGIPLETSAKNATNVQSFMTAAETKRMGPGATAG 151
DB 121 NNCDDLTTRKVVDTYTAKEPADSLGIPLETSAKNATNVQSFMTAAETKRMGPGATAG 179
OY 152 GAESKNVXI-QSTPVKQSGGGCC 173
DB 178 VNNKPTVQVGGQGGQVGSAGGCC 201

```

RESULT 8

Q9P8J7 PRELIMINARY; PRT; 201 AA.

AC Q9P8J7; 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE YPLA.

GN YPLA.

OS Aspergillus awamori.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiiales; Ascomycaceae; mitosporic Trichocomaceae; Aspergillus.

NCBI\_TaxID=105351;

RP SEQUENCE FROM N.A.

RA Salonen M., Wang H., Valkonen M., Vasara T., Huuskonen A., Ward M., Penttilä M.;

RT "the secretory genes ypl1/ypl4 and nas1/nas4 from the filamentous fungi Trichoderma reesei and Aspergillus niger: evidence for global transcriptional regulation of the secretory pathway in Trichoderma reesei";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RM -1-SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

CC EMBL: AF244543; AF6333.1; .

DR HSP: P01112; IPI.

DR Interpro: IPR003579; GTPase\_Rab.

DR Interpro: IPR001806; Ras\_trnsfmg.

DR Interpro: IPR002078; Sig54\_interact.

DR Interpro: IPR005225; Small\_GTP.

DR Pfam: PF00071; Ras; 1.

DR PRINTS: PR00449; RASTRNSFRNG.

DR SMART: SM00175; RAB; 1.

DR TIGRfams: TIGR00231; small\_GTP; 1.

DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.

KW GTP-binding; Lipoprotein.

SO SEQUENCE 201 AA; 22335 MW; 9D710FBAEBDE279 CRC64;

Query Match 73.8%; Score 653.5; DB 3; Length 201;  
Best Local Similarity 67.2%; Pred. No. 5e-45;  
Matches 137; Conservative 12; Mismatches 18; Indels 37; Gaps 3;

QY 4 MNPEYDYLFKLLIGDSGVGRKSCILLRFADDTYESYISTIGVDFKIRITELDKGTIKLQ 63  
|||||  
DB 1 MNPEYDYLFKLLIGDSGVGRKSCILLRFADDTYESYISTIGVDFKIRITELDKGTIKLQ 60

QY 64 I-----ESFNNVKQWLOEIDRYASENNKLLVG 91  
|  
DB 61 IMDTAGGERFRITTSYRYGAHGIYVDYTDQESFNNVKQWLOEIDRYASENNKLLVG 120

QY 92 NKCDLITRKVVDTTAKFADSGIPLETSAKNATNVEOSFMTAAEIKRKMGPGATAG 151  
|||  
DB 121 NMSDEBKVEYETAKFADSLGIPLETSAKNATNVEOAFITMAQIKERMG--TAT 177

QY 152 GAESNNKIOSTPVKQSG--GGCC 173  
|  
DB 178 VNNKPTVQVQGGVGSAGGCC 201

RESULT 9

Q40203 PRELIMINARY; PRT; 202 AA.

AC Q40203; 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE RAB1C.

GN RAB1C.

OS Lotus japonicus.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae; Lotus.

NCBI\_TaxID=34305;

RP SEQUENCE FROM N.A.

RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Baah J.,

RP SEQUENCE FROM N.A.

RC TISSUE=ROOT NODULES; PubMed=9076991;

RX MEDLINE=97231679; Pubmed=9076991;

RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;

RT "Identification of new protein species among 33 different small GTP-binding proteins encoded by cDNAs from Lotus japonicus, and expression of corresponding mRNAs in developing root nodules";

RT of corresponding mRNAs in developing root nodules";

RL Plant J. 11:237-250(1997).

CC -1-SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

DR EMBL: 273932; CA98160.1; .

DR HSP: P05713; 3RAB.

DR Interpro: IPR003579; GTPase\_Rab.

DR Interpro: IPR001806; Ras\_trnsfmg.

DR Interpro: IPR002078; Sig54\_interact.

DR Interpro: IPR005225; Small\_GTP.

DR Pfam: PF00071; Ras; 1.

DR PRINTS: PR00449; RASTRNSFRNG.

DR SMART: SM00175; RAB; 1.

DR TIGRfams: TIGR00231; small\_GTP; 1.

DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.

KW GTP-binding; Lipoprotein.

SO SEQUENCE 202 AA; 22633 MW; 91633EBFAD07662F CRC64;

Query Match 72.7%; Score 644; DB 10; Length 202;  
Best Local Similarity 65.8%; Pred. No. 2.9e-44;  
Matches 133; Conservative 8; Mismatches 27; Indels 34; Gaps 3;

QY 4 MNPEYDYLFKLLIGDSGVGRKSCILLRFADDTYESYISTIGVDFKIRITELDKGTIKLQ 63  
|||||  
DB 1 MNPEYDYLFKLLIGDSGVGRKSCILLRFADDTYESYISTIGVDFKIRITELDKGTIKLQ 60

QY 64 I-----ESFNNVKQWLOEIDRYASENNKLLVG 91  
|  
DB 61 IMDTAGGERFRITTSYRYGAHGIYVDYTDQESFNNVKQWLOEIDRYASENNKLLVG 120

QY 92 NKCDLITRKVVDTTAKFADSGIPLETSAKNATNVEOSFMTAAEIKRKMGPGATAG 151  
|||  
DB 121 NKCDLITRKVVDTTAKFADSGIPLETSAKNATNVEOAFITMAQIKERMG--TAT 177

QY 152 GAESNNKIOSTPVKQSG--GGCC 173  
|  
DB 180 NARPPYQIRGQPVNQR--SGCC 200

RESULT 10

Q9SEH3 PRELIMINARY; PRT; 202 AA.

AC Q9SEH3; 01-MAR-2000 (TREMBLrel. 13, Created)

DT 01-MAR-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE RAS-related small GTP-binding protein (RAS-related small GTP-binding protein RAB1C) (At4G17530/DLA800C).

GN RAB1C OR AT4G17530.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

NCBI\_TaxID=3702;

RP SEQUENCE FROM N.A.

RA Bischoff F., Godde M., Palme J.;

RT "Arabidopsis involved in ER-Golgi traffic in meristematic cells in Arabidopsis";

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RM -1-SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

CC EMBL: 273932; CA98160.1; .

DR HSP: P05713; 3RAB.

DR Interpro: IPR003579; GTPase\_Rab.

DR Interpro: IPR001806; Ras\_trnsfmg.

DR Interpro: IPR002078; Sig54\_interact.

DR Interpro: IPR005225; Small\_GTP.

DR Pfam: PF00071; Ras; 1.

DR PRINTS: PR00449; RASTRNSFRNG.

DR SMART: SM00175; RAB; 1.

DR TIGRfams: TIGR00231; small\_GTP; 1.

DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.

KW GTP-binding; Lipoprotein.

SO SEQUENCE 202 AA; 22633 MW; 91633EBFAD07662F CRC64;

Query Match 72.7%; Score 644; DB 10; Length 202;  
Best Local Similarity 65.8%; Pred. No. 2.9e-44;  
Matches 133; Conservative 8; Mismatches 27; Indels 34; Gaps 3;

QY 4 MNPEYDYLFKLLIGDSGVGRKSCILLRFADDTYESYISTIGVDFKIRITELDKGTIKLQ 63  
|||||  
DB 1 MNPEYDYLFKLLIGDSGVGRKSCILLRFADDTYESYISTIGVDFKIRITELDKGTIKLQ 60

QY 64 I-----ESFNNVKQWLOEIDRYASENNKLLVG 91  
|  
DB 61 IMDTAGGERFRITTSYRYGAHGIYVDYTDQESFNNVKQWLOEIDRYASENNKLLVG 120

QY 92 NKCDLITRKVVDTTAKFADSGIPLETSAKNATNVEOSFMTAAEIKRKMGPGATAG 151  
|||  
DB 121 NKCDLITRKVVDTTAKFADSGIPLETSAKNATNVEOAFITMAQIKERMG--TAT 177

QY 152 GAESNNKIOSTPVKQSG--GGCC 173  
|  
DB 180 NARPPYQIRGQPVNQR--SGCC 200

Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawada J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.,  
 RT Arabidopsis cDNA clones."  
 CC Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 RL -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.  
 DR EMBL: AF127134; AAF22133.1; -  
 DR EMBL: AL161546; CAB78756.1; -  
 DR EMBL: AY052204; AK97675.1; -  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfrmng.  
 DR InterPro: IPR002078; Sig54\_interact.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAMS: TIGR00231; small\_gtp; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding; Lipoprotein.  
 FT VARIANT 70 70 F -> L.  
 SQ SEQUENCE 202 AA; 23318 MW; 954FA24C3110FC12 CRC64;  
 Query Match 72.5%; Score 642; DB 10; Length 202;  
 Best Local Similarity 65.8%; Pred. No. 4.2e-44;  
 Matches 133; Conservative 11; Mismatches 24; Indels 34; Gaps 3;  
 QY 4 NMPEYDYLFKLLIDSGVSKCLLRFPADDTYTESYSTIGVDKIRTIELDKTKIKQ 63  
 DB 1 NMPEYDYLFKLLIDSGVSKCLLRFPADDTYTESYSTIGVDKIRTIELDKTKIKQ 60  
 QY 64 I-----ESFNWKQWLOEIDRYASFNVKLLVG 91  
 DB 61 IMDTAGQERFRTTSSYRGAGHGVYDYVDLSEFNNKQWLOEIDRYASFNVKLLVG 120  
 QY 92 NKCDLTKRKVDYTTAKERFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGATAG 151  
 DB 121 NKCDLTSQKVSTERAKAFADSLGIPLETSAKNATNVEANMATMAIKTKMA-SQPRAG 179  
 QY 152 GAESKNVAKI-QSTPY-KOSGGGCC 173  
 DB 180 GSKPPTVQIRGQPVNQO-SGCC 200  
 RESULT 11  
 Q9HDT5 PRELIMINARY; PRT; 202 AA.  
 AC Q9HDT5;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Small GTP-binding protein YPT1.  
 GN YPT1.  
 OS Trichoderma reesei (Hypocrea jecorina).  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; Hypocreaceae; Hypocrea.  
 OX NCBI\_TaxID=51453;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RUIC-30;  
 RA Saloheimo M.L.A., Mann H., Valkonen M., Vasara T., Huuskonen A.,  
 RA Rikonen M., Pakula T., Ward M., Penttilae M.,  
 RT "The secretory genes ypt1/ypt4 and nsf1/nsf4 from the filamentous  
 RT fungi Trichoderma reesei and Aspergillus niger var. awamori: evidence  
 RT for global transcriptional regulation of the secretory pathway in T.  
 RT reesei."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL: AJ277108; CAC17744.1; -

HSSP: P01112; 1PLJ.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfrmng.  
 DR InterPro: IPR002078; Sig54\_interact.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAMS: TIGR00231; small\_gtp; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 202 AA; 22421 MW; A60E10E2AB12E7 CRC64;  
 Query Match 72.2%; Score 640; DB 3; Length 202;  
 Best Local Similarity 65.7%; Pred. No. 6.1e-44;  
 Matches 134; Conservative 13; Mismatches 21; Indels 36; Gaps 4;  
 QY 4 NMPEYDYLFKLLIDSGVSKCLLRFPADDTYTESYSTIGVDKIRTIELDKTKIKQ 63  
 DB 1 NMPEYDYLFKLLIDSGVSKCLLRFPADDTYTESYSTIGVDKIRTIELDKTKIKQ 60  
 QY 64 I-----ESFNWKQWLOEIDRYASFNVKLLVG 91  
 DB 61 IMDTAGQERFRTTSSYRGAGHGVYDYVDLSEFNNKQWLOEIDRYASFNVKLLVG 120  
 QY 92 NKCDLTKRKVDYTTAKERFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPGATAG 151  
 DB 121 NKSDMSDKRVVYTTAKERFADSLGIPLETSAKNATNVEANMATMAIKTKMA--GTITA 178  
 QY 152 GAESKNVAKI-QSTPY-KOSGGGCC 173  
 DB 179 NNTKPSVHVQCGQGVNSSSSCC 202  
 RESULT 12  
 Q08154 PRELIMINARY; PRT; 203 AA.  
 AC Q08154;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE GTP-binding protein.  
 OS Pisum sativum (Garden pea).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosida I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.  
 OX NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LEAF;  
 RX MEDLINE=94290824; PubMed=8019783;  
 RA Nagano Y., Mural N., Matsuno R., Sasaki Y.,  
 RT "Isolation and characterization of cDNAs that encode eleven small GTP-  
 RT binding protein from Pisum sativum."  
 RL Plant Cell Physiol. 34:447-455(1993).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL: D12549; BAA02117.1; -  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfrmng.  
 DR InterPro: IPR002078; Sig54\_interact.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAMS: TIGR00231; small\_gtp; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 203 AA; 22463 MW; 6F2E1007A31EF3F4 CRC64;  
 Query Match 71.8%; Score 636.5; DB 10; Length 203;  
 Best Local Similarity 64.9%; Pred. No. 1.2e-43;  
 Matches 131; Conservative 11; Mismatches 27; Indels 33; Gaps 2;

```

OY 4 MNPEDYLFKLLIGSGVKSCLLRADDTTSTSTIGVDFKRTIETDGTIKLQ 63
DB 1 MNPEDYLFKLLIGSGVKSCLLRADDTTSTSTIGVDFKRTIETDGTIKLQ 60
OY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91
DB 61 IMDTAGOERRRTTSSYRGAGHIIIVYDVTDESEFNNVKOMLEIDRYASENVKLLVG 120
OY 92 NKCDLTTKRVYDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAEIKRKGPGATAG 151
DB 121 NKSDLTARVYSDYDKFEFADGIGPMEFSAKDATNVEGAFMAAIAIKRMAQSPSAN 180
OY 152 GAESKNVYIOSTFVKOSGGCC 173
DB 181 NARPPYVQIRGQPVNGK -GGCC 201

RESULT 13
ID 08M4S8 PRELIMINARY; PRT: 202 AA.
AC 08M4S8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE AR4917530/414800.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Caninici P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones";
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY060495; AL31108.1; -
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001806; Ras_trnsfmg.
DR InterPro: IPR002078; S1954_interact.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PR00449; RASTRNSFRMG.
DR SMART: SM00175; Rab; 1.
DR PROSITE: PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
DR GTP-binding, Lipoprotein.
SQ SEQUENCE 202 AA; 22286 MW; 954FA12C3110FC12 CRC64;

Query Match 71.6%; Score 634.5; DB 10; Length 202;
Best Local Similarity 65.8%; Pred. No. 1,7e-43;
Matches 131; Conservative 11; Mismatches 24; Indels 33; Gaps 2;
OY 4 MNPEDYLFKLLIGSGVKSCLLRADDTTSTSTIGVDFKRTIETDGTIKLQ 63
DB 1 MNPEDYLFKLLIGSGVKSCLLRADDTTSTSTIGVDFKRTIETDGTIKLQ 60
OY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91
DB 61 IMDTAGOERRRTTSSYRGAGHIIIVYDVTDESEFNNVKOMLEIDRYASENVKLLVG 120
OY 92 NKCDLTTKRVYDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAEIKRKGPGATAG 151
DB 121 NKNDLTSOKVSTETAKAFADSLGIPLETSAKNATNVEAFMAAIAIKRMA -SOPAG 179
OY 152 GAESKNVYIOSTFVKOSGG 170
DB 180 GSKPPYVQIRGQPVNGSG 198

```

```

RESULT 14
ID 09FPJ4 PRELIMINARY; PRT: 202 AA.
AC 09FPJ4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE AT5947200 (Putative ras-related small GTP-binding protein).
GN AT5947200/MOL5.5 OR AT5947200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Shino P., Chao C., Brooks S., Chen H., Cheuk R., Johnson-Hopson C.,
RA Khan S., Kim C.J., Banh J., Bowser L., Chung M.K., Goldsmith A.D.,
RA Jones T., Karlin-Neumann G., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakano H.,
RA Southwick A., Toriumi M., Yamada K., Yu G., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
NCBI_TaxID=3702;
[2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Caninici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shino P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Arabidopsis full length cDNA clones";
RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.
DR EMBL: AF134990; AA60342.1; -
DR EMBL: AY080753; AA185999.1; -
DR HSP: P05713; 3RAB.
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001806; Ras_trnsfmg.
DR InterPro: IPR002078; S1954_interact.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PR00449; RASTRNSFRMG.
DR SMART: SM00175; Rab; 1.
DR TRIGRAMS: TIGR00231; small_GTP_1.
DR PROSITE: PS00675; SIGMA54_INTERACT_1; UNKNOWN_1.
KV GTP-binding, Lipoprotein.
SQ SEQUENCE 202 AA; 22313 MW; D0129A296C3EFDPC CRC64;

Query Match 71.4%; Score 633; DB 10; Length 202;
Best Local Similarity 65.8%; Pred. No. 2,2e-43;
Matches 133; Conservative 10; Mismatches 25; Indels 34; Gaps 3;
OY 4 MNPEDYLFKLLIGSGVKSCLLRADDTTSTSTIGVDFKRTIETDGTIKLQ 63
DB 1 MNPEDYLFKLLIGSGVKSCLLRADDTTSTSTIGVDFKRTIETDGTIKLQ 60
OY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91
DB 61 IMDTAGOERRRTTSSYRGAGHIIIVYDVTDESEFNNVKOMLEIDRYASENVKLLVG 120
OY 92 NKCDLTTKRVYDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAEIKRKGPGATAG 151
DB 121 NKNDLTSOKVSTETAKAFADSLGIPLETSAKNATNVEAFMAAIAIKRMA -SOPAG 179
OY 152 GAESKNVYIOSTFVKOSGGCC 173
DB 180 GAKPPYVQIRGQPVNG -SGCC 200

```

## RESULT 15

024112 ID 024112 PRELIMINARY; PRT; 203 AA.  
 AC 024112;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Small GTP-binding protein.  
 OS Nicotiana plumbaginifolia (Leadwort-leaved tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4092;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Borisjuk N., Sitallo L., Kaydamov C., Senger S., Tewes A.,  
 RA Manteuffel R.;  
 RT "Cloning and characterization of cDNA clones differentially expressed  
 RT during somatic embryogenesis of Nicotiana plumbaginifolia: a mRNA  
 RT differential display approach."  
 RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL: Y08425; CAA69701.1; .  
 DR HSSP: P01112; IPL.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001230; Prenyl\_site.  
 DR InterPro: IPR001806; Ras\_transfmg.  
 DR InterPro: IPR002078; S1954\_interact.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAMS: TIGR00231; small\_GTP; 1.  
 DR PROSITE: PS00294; PRENYLATION; UNKNOWN\_1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW GTP-binding; Lipoprotein.  
 SQ SEQUENCE 203 AA; 22516 MW; 13799234FF719DE7 CRC64;

Query Match 71.4%; Score 632.5; DB 10; Length 203;  
 Best Local Similarity 63.9%; Pred. No. 2.5e-43;

Matches 129; Conservative 12; Mismatches 28; Indels 33; Gaps 2;

QY 4 MNPEYDYLFLKLLIGDSGVGKSCILLRADPTYESYSTIGVDPKRTIELDGKTKLQ 63  
 DB 1 MNPEYDYLFLKLLIGDSGVGKSCILLRADPTYESYSTIGVDPKRTIELDGKTKLQ 60  
 QY 64 I-----ESFNNKQWLOEIDRYASENVNKLIVG 91  
 DB 61 IMDTAGQERFRTISSYRGAHGIIVYDVTQDSFNNVKQWLSRIDRYASDSYKLLVIG 120  
 QY 92 NKCDLTTRKVVYDTTAAEFADSLGIPFLETSAKNATNVEQSFMTAAEIKRKGPGATAG 151  
 DB 121 NKCDLTAAQVSTETAFADEIGIPFLETSAKNATNVEQAFMAAASIKRMAASQPASN 180  
 QY 152 GAESNVKIOSTPYKQSGGCC 173  
 DB 181 NARPPVQIRGQPVNQR-SGCC 201

Search completed: June 18, 2003, 15:23:06  
 Job time : 82 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 15:21:20 ; Search time 27 Seconds  
(without alignments)  
188.525 Million cell updates/sec

Title: US-09-820-003b-2

Sequence: 1 MSSNPXYDLFFLLIGDS.....EKSNNKIQSTPYNQSGGCGC 173

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

1: Issued Patents.AA:\*  
2: /cgn2\_6/prodata/1/aa/58\_COMB pep:\*  
3: /cgn2\_6/prodata/1/aa/58\_COMB pep:\*  
4: /cgn2\_6/prodata/1/aa/58\_COMB pep:\*  
5: /cgn2\_6/prodata/1/aa/PCFUS\_COMB pep:\*  
6: /cgn2\_6/prodata/1/aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	813.5	91.8	201	2	US-08-531-525-13
2	813.5	91.8	201	2	US-08-718-270A-13
3	766.5	86.5	201	4	US-08-916-901-3
4	766.5	86.5	201	4	US-08-154-602-3
5	760.5	85.8	201	4	US-08-916-901-8
6	760.5	85.8	201	4	US-08-154-602-8
7	702.5	79.3	202	2	US-08-531-525-14
8	702.5	79.3	202	2	US-08-718-270A-14
9	439	40.5	215	2	US-08-531-525-10
10	439	40.5	215	2	US-08-718-270A-10
11	424	47.5	207	2	US-08-531-525-35
12	424	47.5	207	2	US-08-718-270A-35
13	417.5	47.1	207	3	US-08-824-873-4
14	417.5	47.1	207	3	US-09-158-184-4
15	417.5	47.1	205	2	US-08-531-525-25
16	417.5	47.1	205	2	US-08-718-270A-25
17	390	44.0	198	2	US-08-531-525-51
18	390	44.0	198	2	US-08-718-270A-51
19	366	41.3	194	2	US-08-531-525-34
20	366	41.3	194	2	US-08-718-270A-34
21	334	37.7	208	2	US-08-531-525-17
22	334	37.7	208	2	US-08-718-270A-17
23	330.5	37.3	212	4	US-09-399-913-67
24	330.5	37.3	212	4	US-08-531-525-19
25	330.5	37.3	218	2	US-08-718-270A-19
26	330	37.2	212	2	US-08-531-525-18
27	330	37.2	212	2	US-08-718-270A-18

28	322	36.3	210	2	US-08-531-525-16	Sequence 16, Appl
29	322	36.3	210	2	US-08-718-270A-16	Sequence 16, Appl
30	319	36.0	214	2	US-08-531-525-52	Sequence 52, Appl
31	319	36.0	214	2	US-08-718-270A-52	Sequence 52, Appl
32	316.5	35.7	190	2	US-08-824-873-3	Sequence 3, Appl
33	316.5	35.7	190	2	US-09-198-184-3	Sequence 3, Appl
34	315.5	35.6	213	2	US-08-531-525-11	Sequence 11, Appl
35	315.5	35.6	213	2	US-08-718-270A-11	Sequence 11, Appl
36	307.5	34.7	203	2	US-08-766-551-8	Sequence 8, Appl
37	298.5	33.7	213	2	US-08-531-525-36	Sequence 36, Appl
38	298.5	33.7	213	2	US-08-718-270A-36	Sequence 36, Appl
39	292.5	33.0	191	4	US-09-075-454-3	Sequence 3, Appl
40	286	32.3	208	2	US-08-531-525-15	Sequence 15, Appl
41	286	32.3	208	2	US-08-718-270A-15	Sequence 15, Appl
42	263.5	29.7	191	2	US-08-531-525-26	Sequence 26, Appl
43	263.5	29.7	191	2	US-08-718-270A-26	Sequence 26, Appl
44	255	28.8	217	2	US-08-773-423-3	Sequence 3, Appl
45	254	28.7	213	2	US-08-773-423-8	Sequence 8, Appl

#### ALIGNMENTS

RESULT 1  
US-08-531-525-13  
; Sequence 13, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 58406831e, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides inhibiting the Oncogenic Action  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Wanner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC Compatible  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-Sep-1995  
; CLASSIFICATION: 350  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fether, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Discopyge ommata  
; US-08-531-525-13  
Query Match: 91.8%; Score 813.5; DB 2; Length 201;  
Best Local Similarity 81.2%; Pred. No. 2.3e-87;

Matches 164: Conservative 3; Mismatches 2; Indels 33; Gaps 2;

QY 4 MNPEYDYLFKLLIGDSGVGKSCLLLRPADDTYTESYSTIGVDPKIRTIELDGKTIKQ 63  
|||||  
Db 1 MNPEYDYLFKLLIGDSGVGKSCLLLRPADDTYTESYSTIGVDPKIRTIELDGKTIKQ 60  
QY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91  
1 IWDTAGOERFRITTSYRGAGHIIYVDYDQESFNNVKOMLOEIDRYASENVKLLVG 120  
Db 61 IWDTAGOERFRITTSYRGAGHIIYVDYDQESFNNVKOMLOEIDRYASENVKLLVG 120  
QY 92 NKCDLTKKVVDDYTTAKFADSLGIPLETSAKNATNVEQSEFMTAAEIKRMGPGATAG 151  
|||||  
Db 121 NKCDLTKKVVDDYTTAKFADSLGIPLETSAKNATNVEQSEFMTAAEIKRMGPGATAG 179  
QY 152 GAESKNVKIQSTPVKSGGCC 173  
|||||  
Db 180 GSEKSNVNIQSTPVKSGGCC 201

## RESULT 2

US-08-718-270A-13  
; Sequence 13, Application US/08718270A  
; Patent No. 5910478  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Plucus, Matthew R.  
; APPLICANT: No. 59104781e, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptidomimetics Inhibiting  
; TITLE OF INVENTION: the Oncogenic Action of P21 Ras  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P. C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718.270A  
; FILING DATE: 20-SEP-1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/531,525  
; FILING DATE: 21-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/004,091  
; FILING DATE: 21-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feibet, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 78-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEtical: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Discopyge ommata  
; US-08-718-270A-13

Query Match 91.8%; Score 813.5; DB 2; Length 201;  
Best Local Similarity 81.2%; Pred. No. 2.3e-87;  
Matches 164: Conservative 3; Mismatches 2; Indels 33; Gaps 2;

QY 4 MNPEYDYLFKLLIGDSGVGKSCLLLRPADDTYTESYSTIGVDPKIRTIELDGKTIKQ 63  
|||||  
Db 1 MNPEYDYLFKLLIGDSGVGKSCLLLRPADDTYTESYSTIGVDPKIRTIELDGKTIKQ 60  
QY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91  
1 IWDTAGOERFRITTSYRGAGHIIYVDYDQESFNNVKOMLOEIDRYASENVKLLVG 120  
Db 61 IWDTAGOERFRITTSYRGAGHIIYVDYDQESFNNVKOMLOEIDRYASENVKLLVG 120  
QY 92 NKCDLTKKVVDDYTTAKFADSLGIPLETSAKNATNVEQSEFMTAAEIKRMGPGATAG 151  
|||||  
Db 121 NKCDLTKKVVDDYTTAKFADSLGIPLETSAKNATNVEQSEFMTAAEIKRMGPGATAG 179  
QY 152 GAESKNVKIQSTPVKSGGCC 173  
|||||  
Db 180 GSEKSNVNIQSTPVKSGGCC 201

## RESULT 3

US-08-916-901-3  
; Sequence 3, Application US/08916901  
; Patent No. 5892012  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Puvli  
; TITLE OF INVENTION: RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/916.901  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0367 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-845-4166  
; TELEFAX: 415-855-0555  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: LIVERUT04  
; CLONE: 2514506  
; US-08-916-901-3

Query Match 86.5%; Score 766.5; DB 2; Length 201;  
Best Local Similarity 77.2%; Pred. No. 7.3e-82;  
Matches 156: Conservative 6; Mismatches 7; Indels 33; Gaps 2;

QY 4 MNPEYDYLFKLLIGDSGVGKSCLLLRPADDTYTESYSTIGVDPKIRTIELDGKTIKQ 63



Db 1 MNPEDYLFKLLIGDSGVGKSCLLRFPADDTYESYSTIGVDFKRTIELDGTIKLQ 60  
QY 64 I-----ESFNNVKOMLQIEDRYASENVNKLKLV 91  
Db 61 IMDTAGOERFRTTSSYRGAGIIVYDVTDOESYANVKOMLQIEDRYASENVNKLKLV 120  
QY 92 NKCDLTKRKVDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAAEIKRMGPATAG 151  
Db 121 NKSDLTTRKVDNTAKFADSLGIPLETSAKNATNVEOAFMTMAAEIKRMGPAGAA 180  
QY 152 GAERKNVKIOSTPVKSGGCC 173  
Db 181 G-ERPRLKIDSTPVKPGAGGCC 201

## RESULT 4

US-09-154-602-3  
; Sequence 3, Application US/09154602  
; Patent No. 6300472  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Puri  
; TITLE OF INVENTION: RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/154,602  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/916,901  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0367 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: L1VNTUT04  
; CLONE: 2514506  
; US-09-154-602-3

Query Match 86.5%; Score 766.5; DB 4; Length 201;  
Best Local Similarity 77.2%; Pred. No. 7.3e-82;  
Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;  
QY 4 MNPEDYLFKLLIGDSGVGKSCLLRFPADDTYESYSTIGVDFKRTIELDGTIKLQ 63  
Db 1 MNPEDYLFKLLIGDSGVGKSCLLRFPADDTYESYSTIGVDFKRTIELDGTIKLQ 60  
QY 64 I-----ESFNNVKOMLQIEDRYASENVNKLKLV 91  
||:|||||

Db 61 IMDTAGOERFRTTSSYRGAGIIVYDVTDOESYANVKOMLQIEDRYASENVNKLKLV 120  
QY 92 NKCDLTKRKVDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAAEIKRMGPATAG 151  
Db 121 NKSDLTTRKVDNTAKFADSLGIPLETSAKNATNVEOAFMTMAAEIKRMGPAGAA 180  
QY 152 GAERKNVKIOSTPVKSGGCC 173  
Db 181 G-ERPRLKIDSTPVKPGAGGCC 201

## RESULT 5

US-08-916-901-8  
; Sequence 8, Application US/08916901  
; Patent No. 5892012  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Puri  
; TITLE OF INVENTION: RAB PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/916,901  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0367 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 57006  
; US-08-916-901-8

Query Match 85.8%; Score 760.5; DB 2; Length 201;  
Best Local Similarity 76.2%; Pred. No. 3.7e-81;  
Matches 154; Conservative 7; Mismatches 8; Indels 33; Gaps 2;  
QY 4 MNPEDYLFKLLIGDSGVGKSCLLRFPADDTYESYSTIGVDFKRTIELDGTIKLQ 63  
Db 1 MNPEDYLFKLLIGDSGVGKSCLLRFPADDTYESYSTIGVDFKRTIELDGTIKLQ 60  
QY 64 I-----ESFNNVKOMLQIEDRYASENVNKLKLV 91  
||:|||||  
Db 61 IMDTAGOERFRTTSSYRGAGIIVYDVTDOESYANVKOMLQIEDRYASENVNKLKLV 120  
QY 92 NKCDLTKRKVDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAAEIKRMGPATAG 151  
Db 121 NKSDLTTRKVDNTAKFADSLGIPLETSAKNATNVEOAFMTMAAEIKRMGPAGAA 180

```

QY      152 GAESNVKIQSTPVKQSGGGCC 173
          | : | | | | | : | | |
Db      181 G-ERNPKIDSTPVKSASGGCC 201

```

**RESULT 6**  
**RE-09-1E**

```

US-09-154-602-8
Sequence 8 , Application US/09154602
Patent No. 6300472
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: RAB PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3114 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,602
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,901
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ. ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 57006
US-09-154-602-8

```

Query Match	85.8%	Score 760.5;	DB 4,	Length 201;
Best Local Similarity	76.2%	Pred. No. 3.7e-81;		
Matches 154; Conservative	7;	Mismatches 8;	Indels 33;	Gaps 2

```

QY      4 MNEPYTLKLLIGSGKSCLLLRADDTYTESYSLTGVPKRTIELDQTKRLQ 63
Db      1 MNEPYTLKLLIGSGVKSCLLRADDTYTESYSLTGVPKRTIELDQTKRLQ 60
QY      64 I-----ESFNNVKQWLQEIIDRYASENNKLLVG 91
Db      61 IMDTAGGERPRTVTSYSGKANGAILVYDPTYDQESYANVKQWLQEIIDRYASENNKLLVG 120
QY      92 NKCDLTKKRVDDTTAKEPADSLGIPLETSAKNNTNVEGSPMTAAEIKRRMGPGATAG 151
Db      121 NNSDLTKKRVDDTTAKEPADSLGVPLETSAKNNTNVEQAQFMTAAEIKRRMGPGAAG 180
QY      152 GAEKSNVYKIQSTPVKQSGGCC 173
Db      181 G-ERPMUKIDSTPVKASGGCC 201

```

RESULT 7  
US-08-531-525-14  
; Sequence 14, Application US/08531525

Query Match	79.3%	Score 702.5	DB 2	Length 202;
Best Local Similarity	72.2%	Pred. No. 2.3e-74;		
Matches 148; Conservative	9;	Mismatches 13;	Indels 35;	Gaps 4;

QY	1	MSANNEVYLERKLLIGSGVGRKCLLRFADPTVYSYSTGVDFKRTIELDOKTI	60
Db	1	MSANNPDIYLERKLLIGSDGVGRKCLLRFADPTVYSYSTGVDFKRTIELDOKTI	60
QY	61	KLQI-----ESFNNVKQMLDEIDRYASENNKL	88
Db	61	KLQIWDTAGQERFRITTSYRKGAGHIIWVYDWDQSEFNNVKQMLDEIDRYASENNKL	120
QY	89	LVGNKCDLTKKRVYDTTAKERFADSLGIPLETSAKNAITNVEGSFMTMAAELIKRNGPGA	148
Db	121	LVGNNSDLTKRK-VDTTAKERFADSLGIPLETSAKNAITNVEQAFMTMAAELIKRNGP-I	178
QY	149	TAGGAESNVKTIQSTPVKQSGGGCC	173
Db	179	TASDS-KPSVKIINSSTPSHKNGGCC	202

RESULT 8  
US-08-718-270A-14  
; Sequence 14, Application US/08718270A

Patent No. 5910478.  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 59104781e, John F.  
APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptidomimetics Inhibiting  
the Oncogenic Action of P21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/08/718, 270A  
APPLICATION NUMBER: US/08/718, 270A  
CLASSIFICATION: 514  
PRIORITY APPLICATION DATA:  
PRIORITY APPLICATION NUMBER: US 08/531,525  
FILING DATE: 21-SEP-1995  
PRIORITY APPLICATION DATA:  
PRIORITY APPLICATION NUMBER: US 60/004,091  
FILING DATE: 21-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Feiber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 78-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 202 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Lymnea stagnalis  
US-08-718-270A-14

Query Match  
Best Local Similarity 79.3%; Score 702.5; DB 2; Length 202;  
Matches 148; Conservative 9; Mismatches 13; Indels 35; Gaps 4;

QY 1 MESNPEVDYLFKLLIGDSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDKGTI 60  
DB 1 MGNMPDDYLFKLLIGDSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDKGTI 60  
QY 61 KIQI-----ESFNKQWLOEIDRVASENVNKL 88  
DB 61 KQIWDVAGQERFRTTSSYRGAGIIVYDVTDESSENNKQWLOEIDRVASENVNKL 120  
QY 89 LVGNCDLTTKRVVDYTTAKERPADSLGIPLETSAKNATNVEOSFTWMAEIKRMGPGA 148  
DB 121 LVGNNSDLYTTER-VDTTAKERPADSLGIPLETSAKNATNVEOAFWMAEIKRMGP-I 178  
QY 149 TAGGAESKNVKSIOSTPVKSGGCC 173  
DB 179 TASDS-KPSVKNSTPSANKGCC 202

RESULT 9

US-08-531-525-10  
Sequence 10, Application US/08531525  
Patent No. 5840683  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 58406831e, John F.  
APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
of P21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/08/531,525  
APPLICATION NUMBER: US/08/531,525  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
PRIORITY APPLICATION NUMBER: 530  
FILING DATE: 21-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Feiber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 37-94  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 215 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Arabidopsis thaliana  
US-08-531-525-10

Query Match  
Best Local Similarity 49.5%; Score 439; DB 2; Length 215;  
Matches 93; Conservative 28; Mismatches 44; Indels 40; Gaps 4;

QY 7 EYDLFLLKLLIGDSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDKGTI 64  
DB 11 DYDLFLLKLLIGDSGVKSCLLRFPADDTYESYSTIGVDFKIRITELDKGTI 70  
QY 65 -----ESFNKQWLOEIDRVASENVNKL 95  
DB 71 TAGQERTTTAYRGAGIIVYDVTDESSENNKQWLOEIDRVASENVNKL 130  
QY 96 L-TTKRVVDYTTAKERPADSLGIPLETSAKNATNVEOSFTWMAEIKRMGPGA 154  
DB 131 MDESKRAVFPKAGQALADEYGIKFFETSATKMLNVEEVEFISGRDIKRLSD--TDSRAE 188  
QY 155 KSNVKSIOSTPVKSGGCC-----GCC 173  
DB 189 PATIKISQTDQAGAGQATKRSACC 213

RESULT 10  
US-08-718-270A-10  
Sequence 10, Application US/08718270A  
Patent No. 5910478  
GENERAL INFORMATION:

APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 59104781e, John F.  
APPLICANT: Adajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptidomimetics Inhibiting  
TITLE OF INVENTION: the Oncogenic Action of p21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/718,270A  
FILING DATE: 20-SEP-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/531,525  
FILING DATE: 21-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/004,091  
FILING DATE: 21-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 78-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 215 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Arabidopsis thaliana  
US-08-718-270A-10

Query Match 49.5%; Score 439; DB 2; Length 215;  
Best Local Similarity 45.4%; Pred. No. 2e-43;  
Matches 93; Conservative 28; Mismatches 44; Indels 40; Gaps 4;  
QY 7 EYDYFKLLIGDSGVGKSCLLRFADDTYESYSTIGVDKRTIELDGKTIKQI-- 64  
DB 11 DYDIYKLLIGDSGVGKSCLLRFSDFTTSFTTIGIDKRTIELDGKRIKQIWD 70  
QY 65 -----ESNNVKOMLOEIDRYASENNKLLVGNKCD 95  
DB 71 TAGOERFTTAYYNGANGILLVYDTDESSFNRRNIRNIEQASDNVKNILVGNKRD 130  
QY 96 L-TTKKVVDTAAKEFADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPGATAGAAE 154  
DB 131 MDESRAVPTAKGALADEYGIKFPETSAKTLNVEEFESIGRIDIKQLSD--TDSRAE 188  
QY 155 KSNVYIOSTPVKQSG-----GCC 173  
DB 189 PATIKISQTDAAAGAGATQKRSACC 213

RESULT 11  
US-08-531-525-35  
; Sequence 35, Application US/08531525

Patent No. 5840683  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 58406831e, John F.  
APPLICANT: Adajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
TITLE OF INVENTION: of p21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/531,525  
FILING DATE: 21-SEP-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 37-94  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8089  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 207 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Discopyge ommata  
US-08-531-525-35

Query Match 47.9%; Score 424; DB 2; Length 207;  
Best Local Similarity 42.9%; Pred. No. 1.1e-41;  
Matches 82; Conservative 38; Mismatches 41; Indels 30; Gaps 1;  
QY 8 YDYLFKLLIGDSGVGKSCLLRFADDTYESYSTIGVDKRTIELDGKTIKQI-- 64  
DB 4 YDYLFKLLIGDSGVGKSCLLRFSEDAFNTFTSTIGIDFRIVEIDGKRIKQIWD 63  
QY 65 -----ESNNVKOMLOEIDRYASENNKLLVGNKCD 97  
DB 64 AGOERFTTAYYNGANGIMKVDITNESFPDKIRNIRIEHASSDVERMLKGMCDMN 123  
QY 98 TKKVVDTAAKEFADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPGATAGAEKSN 157  
DB 124 EKROYSKERGEKLADYGIKFPETSAKSSINVEEAFITLARDIMTKLNKNNENSLQAV 183  
QY 158 VKIOSTPVKQSG-----GCC 173  
DB 184 DKIKSPKPKPS 194

RESULT 12  
US-08-718-270A-35  
; Sequence 35, Application US/08718270A  
; Patent No. 5910478  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.



STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/198,184  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/824,873  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0240 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 207 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
US-09-198-184-4

Query Match 47.2% Score 418; DB 3; Length 207;  
Best Local Similarity 42.9%; Pred. No. 5.5e-41;  
Matches 84; Conservative 33; Mismatches 47; Indels 33; Gaps 1;

QY 4 MNPEYDLEFKLLIGDSGVGKSCLLRFADDTYTESYISTIGVDFKRTIELDGKTIKQ 63  
DB 1 MAKTYDLEFKLLIGDSGVGKTCVLFREFSDAFNSTFISTIGDFKRTIELDGKRIKQ 60  
QY 64 I-----ESFNNVKOMLOEIDRYASENVKLLVG 91  
DB 61 IWDTAGOERFTITAYRANGIMLVYDITNEKSPDNIRNMINIEHSAADVERKILG 120  
QY 92 NKCDLTTRKVVDTTAKFADSLGIPLETSAKNATNVEOSFTMAAEIKRKGPGATAG 151  
DB 121 NKCDVNDKRVSKERGERKIALDYGIKFMETSAKANINVENAFITLARDIKAKDKLEGN 180  
QY 152 GAESNVKIQSTPVKQ 167  
DB 181 SPOGSGVGYKTPDQ 196

RESULT 15  
US-08-531-525-25  
Sequence 25, Application US/08531525  
Patent No. 5840683  
GENERAL INFORMATION:  
APPLICANT: Hiavva, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 5840683le, John F.  
APPLICANT: Adajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
TITLE OF INVENTION: of P21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder

STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/531,525  
FILING DATE: 21-SEP-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Feiber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 37-94  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Canis familiaris  
US-08-531-525-25

Query Match 47.1% Score 417.5; DB 2; Length 205;  
Best Local Similarity 43.5%; Pred. No. 6.2e-41;  
Matches 83; Conservative 33; Mismatches 44; Indels 31; Gaps 1;

QY 8 YVILFRLLIGDSGVGKSCLLRFADDTYTESYISTIGVDFKRTIELDGKTIKQ--- 64  
DB 4 YVILFRLLIGDSGVGKTCVLFREFSDAFNSTFISTIGDFKRTIELDGKRIKQIMDT 63  
QY 65 -----ESFNNVKOMLOEIDRYASENVKLLVGKCDL 96  
DB 64 AGOERFTITAYRANGIMLVYDITNEKSPDNIRNMINIEHSAADVERKILG 123  
QY 97 TTRKVVDTTAKFADSLGIPLETSAKNATNVEOSFTMAAEIKRKGPGATAGAEKS 156  
DB 124 NDKRVSKERGERKIALDYGIKFMETSAKANINVENAFITLARDIKAKDKLEGNSPOGS 183  
QY 157 NVKIQSTPVKQ 167  
DB 184 NQGVKTPDQ 194

Search completed: June 18, 2003, 15:25:01  
Job time : 28 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 15:20:05 ; Search time 28 Seconds  
(without alignments)  
668,564 Million cell updates/sec

Title: US-09-820-003b-2  
Perfect score: 886  
Sequence: 1 MSSMNPEDYLFKLLIGDSGVCSCLLRFADDTYESYISPIGVDRKIRITELDKGTI 173

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues  
Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

# Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEM\_PUB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEM\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEM\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEM\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEM\_PUB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEM\_PUB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEM\_PUB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	886	100.0	173	10 US-09-820-003A-2	Sequence 2, Appl1
2	860	97.1	222	10 US-09-820-003A-4	Sequence 4, Appl1
3	768.5	86.7	224	9 US-10-102-806-466	Sequence 466, App
4	766.5	86.5	201	10 US-09-967-736-3	Sequence 3, Appl1
5	760.5	85.8	201	10 US-09-967-736-8	Sequence 8, Appl1
6	444	46.7	207	10 US-09-794-257-8	Sequence 5, Appl1
7	433	46.6	190	10 US-09-822-860-5	Sequence 5, Appl1
8	403.5	45.5	218	10 US-09-925-300-1571	Sequence 1171, Ap
9	390.5	44.1	246	10 US-09-834-765-766	Sequence 534, App
10	371	41.9	162	10 US-09-794-257-16	Sequence 766, App
11	362.5	40.9	198	10 US-09-945-173-5	Sequence 5, Appl1
12	362.5	40.9	198	10 US-09-972-529-4	Sequence 4, Appl1
13	362.5	40.9	198	10 US-09-972-529-4	Sequence 4, Appl1
14	333	36.3	212	10 US-09-817-198A-4	Sequence 67, Appl1
15	332	36.3	212	10 US-09-817-198A-4	Sequence 1117, Ap
16	322	36.3	212	10 US-09-817-198A-2	Sequence 21, Appl1
17	321	36.2	212	10 US-09-817-198A-2	Sequence 701, App
18	321	36.2	212	10 US-09-817-198A-2	Sequence 4, Appl1
19	315.5	35.6	223	10 US-09-817-199A-4	Sequence 4, Appl1

20	314.5	35.5	223	10 US-09-817-199A-2	Sequence 2, Appl1
21	313	35.3	146	9 US-09-764-868-698	Sequence 698, App
22	312.5	35.3	222	9 US-09-764-868-1106	Sequence 1106, Ap
23	312.5	35.3	226	9 US-09-764-868-684	Sequence 684, App
24	310	35.0	218	10 US-09-817-198A-5	Sequence 5, Appl1
25	306	34.5	188	9 US-09-764-868-1120	Sequence 1120, Ap
26	292.5	33.0	191	10 US-09-794-257-14	Sequence 14, Appl1
27	292.5	33.0	191	12 US-10-051-986-3	Sequence 3, Appl1
28	292.5	33.0	307	9 US-09-764-868-1100	Sequence 1100, Ap
29	292.5	33.0	312	10 US-09-925-302-783	Sequence 783, App
30	292	33.0	59	10 US-09-864-761-35038	Sequence 35038, A
31	267	30.1	216	10 US-09-943-173-10	Sequence 10, Appl1
32	267	30.1	222	9 US-09-764-868-1112	Sequence 1112, Ap
33	267	30.1	225	9 US-09-764-868-692	Sequence 692, Appl1
34	258	29.1	217	9 US-10-036-542-89	Sequence 89, Appl1
35	258	29.1	239	10 US-09-925-301-1077	Sequence 1077, Ap
36	256	28.9	213	9 US-10-036-542-64	Sequence 64, Appl1
37	255	28.8	217	10 US-09-988-974-3	Sequence 3, Appl1
38	254	28.7	213	10 US-09-794-257-5	Sequence 5, Appl1
39	254	28.7	213	10 US-09-988-974-8	Sequence 8, Appl1
40	237.5	26.8	208	9 US-10-108-605-45	Sequence 45, Appl1
41	235.5	26.6	832	10 US-09-834-765-2	Sequence 2, Appl1
42	235	26.5	217	10 US-09-925-300-1364	Sequence 1364, Ap
43	227	25.6	255	9 US-09-764-868-686	Sequence 686, App
44	219.5	24.8	183	10 US-09-765-298A-26	Sequence 26, Appl1
45	219.5	24.8	183	10 US-09-765-298A-26	Sequence 28, Appl1

## ALIGNMENTS

RESULT 1  
US-09-820-003A-2  
; Sequence 2, Application US/09820003A  
; Patent No. US20020142382A1  
; GENERAL INFORMATION:  
; APPLICANT: MERRUOV, Genady et al.  
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
; FILE REFERENCE: C1001196  
; CURRENT FILING DATE: 2001-03-29  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 173  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-820-003A-2

Query Match 100.0%; Score 886; DB 10; Length 173;  
Best Local Similarity 100.0%; Pred. No. 1.2e-76;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSSMNPEDYLFKLLIGDSGVCSCLLRFADDTYESYISPIGVDRKIRITELDKGTI 60  
DB 1 MSSMNPEDYLFKLLIGDSGVCSCLLRFADDTYESYISPIGVDRKIRITELDKGTI 60  
QY 61 KLOESFNVVKQWLEIDRASENVKILVGNKCDLTTRKYVDYTTAREFADSGIPFE 120  
DB 61 KLOESFNVVKQWLEIDRASENVKILVGNKCDLTTRKYVDYTTAREFADSGIPFE 120  
QY 121 TSAKNNTNVEOSFMTMAAEIKRMGAGTAGAKSNVRIQSPVKQSGGCC 173  
DB 121 TSAKNNTNVEOSFMTMAAEIKRMGAGTAGAKSNVRIQSPVKQSGGCC 173  
RESULT 2  
US-09-820-003A-4  
; Sequence 4, Application US/09820003A  
; Patent No. US20020142382A1  
; GENERAL INFORMATION:

APPLICANT: MERKULOV, Gennady et al.  
TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
FILE REFERENCE: C1001196  
CURRENT APPLICATION NUMBER: US/09/820,003A  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 222  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-820-003A-4

Query Match 97.1%; Score 860; DB 10; Length 222;  
Best Local Similarity 84.4%; Pred. No. 5e-74; 0; Indels 32; Gaps 1;

Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MSSNPEYDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGKTI 60  
DB 18 MSSNPEYDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGKTI 77  
QY 61 KIQI-----ESFNNVKQWLQELIDRYASENVKTL 88  
DB 78 KIQIDTAGQERFRTITSSYRGAGIIVYDVTDOESFNNVKQWLQELIDRYASENVKTL 137  
QY 89 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKRMGPAA 148  
DB 138 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKRMGPAA 197  
QY 149 TAGGAESNVKIQSTPVKOSGGGCC 173  
DB 198 TAGGAESNVKIQSTPVKOSGGGCC 222

## RESULT 3

US-10-102-806-466  
Sequence 466, Application US/10102806  
Publication No. US20030054421A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA103PICI  
CURRENT APPLICATION NUMBER: US/10/102,806  
CURRENT FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: 09/925,298  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05881  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 846  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 466  
LENGTH: 224  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-102-806-466

Query Match 86.7%; Score 768.5; DB 9; Length 224;  
Best Local Similarity 76.5%; Pred. No. 2.5e-65;

Matches 156; Conservative 8; Mismatches 7; Indels 33; Gaps 2;

QY 2 SSANPEYDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGKTI 61  
DB 22 SSANPEYDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGKTI 81  
QY 62 LQI-----ESFNNVKQWLQELIDRYASENVKTL 89  
DB 82 LQIMDTAGQERFRTITSSYRGAGIIVYDVTDOESYANVKQWLQELIDRYASENVKTL 141  
QY 90 VGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKRMGPAA 149

DB 142 VGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKRMGPAA 201  
QY 150 AGGAESNVKIQSTPVKOSGGGCC 173  
DB 202 SGG-ERPMLKIDSTPVKPGGGCC 224

## RESULT 4

US-09-967-736-3  
Sequence 3, Application US/09967736  
Patent No. US20020103340A1  
GENERAL INFORMATION:  
APPLICANT: Hallman, Jennifer L.  
Ial, Preeti  
Corley, Neil C.  
Shah, Purni  
TITLE OF INVENTION: RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESS: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/967,736  
FILING DATE: 28-Sep-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/154,602  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0367 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 201 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: L19RTU04

CLONE: 2514506

SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-967-736-3

Query Match 86.5%; Score 766.5; DB 10; Length 201;  
Best Local Similarity 77.2%; Pred. No. 3.3e-65;

Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

QY 4 MNPEYDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGKTI 63  
DB 1 MNPEYDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRTIELDGKTI 60  
QY 64 I-----ESFNNVKQWLQELIDRYASENVKTL 91  
DB 61 IMDTAGQERFRTITSSYRGAGIIVYDVTDOESYANVKQWLQELIDRYASENVKTL 120  
QY 92 NKCDLTTKKVVDYTTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKRMGPAA 151  
DB 121 NKSDLTTKKVVDNTAKFADSLGIPLETSAKNATNVQSFMTMAAEIKRMGPAA 180  
QY 152 GAESNVKIQSTPVKOSGGGCC 173



Mon Jun 23 16:54:10 2003

us-09-820-003b-2.rapb

Page 3

DB 181 G-ERNKLTSTPYKPSAGGCC 201

RESULT 5  
US-09-967-736-8  
Sequence 8, Application US/09967736  
Patent No. US20020103340A1  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
Corley, Neil C.  
Shah, Puryl  
TITLE OF INVENTION: RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESS: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: 28-Sep-2001  
APPLICATION NUMBER: US/09/967,736  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/154,602  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0367 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 57006  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-967-736-8

Query Match 85.8%; Score 760.5; DB 10; Length 201;  
Best Local Similarity 76.2%; Pred. No. 1,2e-64;  
Matches 154; Conservative 7; Mismatches 8; Indels 33; Gaps 2;

DB 61 IMDTAGGERPRITTSYTKAGAGIIVYDVTQESYAVKQWLEIDRYASENVKRLIYG 91  
DB 92 NNCDDLTRKYVDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAEIKRMGATAG 151  
DB 121 NMSDITTKYVDNTAKFADSLGIPLETSAKNATNVEOAFMTMAEIKRMGPGASG 180  
DB 152 GAEKSNVIOSTPYKOSGGCC 173  
DB 181 G-ERNKLTSTPYKPSAGGCC 201

US-09-794-257-8  
Sequence 8, Application US/09794257  
Patent No. US2002009804A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 32702, 23224, 32700, 32712, No. US2002009804A1el  
FILE REFERENCE: 35800, 35895  
CURRENT APPLICATION NUMBER: US/09/794,257  
CURRENT FILING DATE: 2001-02-27  
PRIOR APPLICATION DATA: 2000-02-29  
SOFTWARE: FastSeq for Windows Version 4.0  
NUMBER OF SEQ ID NOS: 16  
SEQ ID NO 8  
LENGTH: 207  
TOPOLOGY: PRT  
ORGANISM: homo sapiens  
US-09-794-257-8

Query Match 46.7%; Score 414; DB 10; Length 207;  
Best Local Similarity 44.2%; Pred. No. 1e-31;  
Matches 84; Conservative 31; Mismatches 37; Indels 38; Gaps 2;

DB 4 MNPETDYLFLILIGDSGVKSCILLRFADDTYESTYSTIGVDKIRITIEDGKTIKQ 63  
DB 1 MAKTYDYLFLILIGDSGVKTCILFFESDAFNTFTSTIGIDKIRITIEDGKTIKQ 60  
DB 64 I-----ESFNVKQWLEIDRYASENVKRLIYG 91  
DB 61 IMDTAGGERPRITTTAYKAGAGIMLVYDITNEKSFDNKIMIRINEHASSDVERMILG 120  
DB 92 NNCDDLTRKYVDYTTAKFADSLGIPLETSAKNATNVEOSFMTMAEIKRM-----G 145  
DB 121 NNCDDMDKQVSKERKELADYGIKLETSAKSNAYEAFITLADITMRKNDMS 180  
DB 146 PGATAGCAEK 155  
DB 181 NSAGAGGPVK 190

RESULT 7  
US-09-822-860-5  
Sequence 5, Application US/09822860  
Patent No. US20020146795A1  
GENERAL INFORMATION:  
APPLICANT: ZHU, ShiaoPing et al.  
TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING "THESE HUMAN RAS-LIKE  
FILE REFERENCE: CL001214  
CURRENT APPLICATION NUMBER: US/09/822,860  
CURRENT FILING DATE: 2001-04-02  
SOFTWARE: FastSeq for Windows Version 4.0  
NUMBER OF SEQ ID NOS: 5  
SEQ ID NO 5  
LENGTH: 190  
TYPE: PRT  
ORGANISM: Discopyge ommata  
US-09-822-860-5

Query Match 46.6%; Score 413; DB 10; Length 190;  
Best Local Similarity 42.1%; Pred. No. 1.1e-31;  
Matches 80; Conservative 38; Mismatches 40; Indels 32; Gaps 1;

DB 9 DYKFLILIGDSGVKSCILLRFADDTYESTYSTIGVDKIRITIEDGKTIKQ---- 64  
DB 1 DYKFLILIGDSGVKTCILFFESDAFNTFTSTIGIDKIRITIEDGKTIKQIWDRA 60  
DB 65 -----ESFNVKQWLEIDRYASENVKRLIYGNCDL 96  
DB 61 GQEFRTITAYKAGAGIMKIVYDITNEKSFDNKIMIRINEHASSDVERMILGNCDL 120

```
QY    97   TTTKVVDDYTAAAEADSLGIPLETSAKNANVEQSPMTMAAETIKRNGPQATGGAEKS 156
      : : : | | | | | | : | : : : : : : : : : : : : : : : : :
Db    121 NERQVSKRGKERGLAIDYCIRKPLETSSAKSSINVEEAFTTLARDIMTKLNKMNNENSLQEA 180

QY    157 NVKIOSTPYK 166
      : : : | |
Db    181 VDKLAKSPPK 190
```

RESULT 8

```

Sequence 1571, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1571
LENGTH: 218
TYPE: prnt
ORGANISM: Homo sapiens
US-09-925-300-1571

```

Query Match	45.58;	Score 403.5;	DB 10;	Length 218;
Best Local Similarity	39.78;	Pred. No. 1.1e-30;		
Matches 85;	Conservative 32;	Mismatches 30;	Indels 67;	Gaps 4

```

QY      8 YDLFFILLIGSGVSKSCLLLRFPADDPVTSYSTGVDRKRTIELDQKTYKLOI --- 64
Db      24 YDLFFILLIGSGVGKTCVLRFSDDAFNTFTSTIGIDRKITVELDQKRIKLOIWT 83.
QY      65 -----ESFNWYKOWLOEIDRYASENKKLVGKCD 95
Db      84 AGGERHITTSYVGANGIMLVYDITNGKSEPNISKLRNIDHANEDEVRMLTGKCD 143
QY      96 LTTKKAVDYTTAKERPADSLGIPLETSAKNAATNVEQSFMTAADIKRMGPGATGAER 155
Db      144 MODKRVVPGKGEJOIAREHGINRFETTSAKANINIEKAVLLTAEDI --- 188
QY      156 SNVKIOSTPYVK-----SGGG-----CC 173
Db      189 ----LTKTPVKEPSENVDISSGGGVWTGSKKCC 218

```

## RESULT 9

Sequence 534, Application US/09925302  
Patent No. US20020044941A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925,302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 534  
LENGTH: 246  
TYPE: PRT

ORGANISM: Homo sapiens  
US-09-925-302-534

Query Match	44.1%;	Score 390.5;	DB 10;	Length 246;
Best Local Similarity	41.7%;	Pred. No. 2.2e-29;		
Matches 78;	Conservative 35;	Mismatches 41;	Indels 33;	Gaps 2;

```

QY      1 MSMSNEEYVLEKRLILIGSGVGKSCILLRAADVPYFSYSTGVPFKRTLELOKRTI 60
      2 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      41 VSMMAAAYDHLKRLILIGSGVGTCLIRRAEDNFENNTYSTGIDCFKRTIYDIEKRTI 100
      56 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      61 KIQI-----ESFNNVKQMLQEIDRYASENVNKL 88
      76 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      101 KIQVQFAGQERKFTTAYVRGAMGILVYDIDKEKSFENIQWMSIKENASAGVERL 160
      126 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      89 LVNKKDDLTTKKAYDITTKAFADSLGIPLETSSKNAATNVEGFSFMTMAELIKRMGPGA 148
      104 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      161 LIGNCKDMQAKRKYQEQADKLARHGIREFETSAKSSMNVDEAFSSILARDIILKSG-GR 219
      176 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      149 TAGAER 155
      164 : : |
Db      220 RSGNGNK 226

```

RESULT 10

```

US-09-834-765-765
Sequence 766, Application US/09834765
Patent No. US20020055478A1
GENERAL INFORMATION:
APPLICANT: Mary Paris
APPLICANT: Pia M. Challita-Eid
APPLICANT: Arthur B. Raitano
APPLICANT: Steve Chappell Mitchell
APPLICANT: Daniel E. H. Afar
APPLICANT: Ava Jakobovits
TITLE OF INVENTION: GMP-BINDING PROTEIN USEFUL IN TREATMENT
TITLE OF INVENTION: AND DETECTION OF CANCER
FILE REFERENCE: 129.60SU1
CURRENT APPLICATION NUMBER: US/09/834,765
CURRENT FILING DATE: 2001-09-21
PRIOR APPLICATION NUMBER: 60/197,647
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 770
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 766
LENGTH: 162
TYPE: PR1
ORGANISM: Homo sapiens
US-09-834-765-766

```

Query Match	41.98;	Score 371;	DB 10;	Length 162;
Best Local Similarity	45.18;	Pred. No. 9e-28;		
Matches 73; Conservative	28;	Mismatches 29;	Indels 32;	Gaps 1;

```

QY 12 FKLIIIGDSGVKSCLLAEADDTYAESITIGVDFKRTIELDGKIKIKQI----- 64
      ||||| ||||| ||| ||| : : : ||||| ||||| |||||
Db 1 FKLIIIGDSGVKTCVLFRRSEDAFESTISIGIDFKRTIELDGKIKIKIWDTAGQE 60
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 -----ESFNVAQWIOEIDRYASEVNNLVLGNKCDLTK 99
      ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 61 REPTITTAAYRCAMGIMLVYDITNKESSFNINMWIRNIEHSAADVERMILGNCDVDK 120
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 100 KVVYDTTAAKEADSLGPLETSAKNATNVEDSFMTMAETK 141
      || : || ||||| ||||| ||||| ||||| |||||
Db 121 ROVSKERGEKLADYIGIKMETSAKANINVENAFTLLADIK 162

```

## RESULT 11

US-09-794-257-16  
; Sequence 16, Application: US/09794257  
; Patent No. US20020009804A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel

```

: TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, NO. US20020009604A1el
: FILE REFERENCE: Human G-Proteins
: CURRENT APPLICATION NUMBER: US/09/794,257
: PRIOR FILING DATE: 2001-02-27
: PRIOR APPLICATION NUMBER: 60/185,606
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 16
: LENGTH: 198
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Pfam accession number PF00071
US-09-794-257-16

```

```

Query Match      40.9%; Score 362.5; DB 10; Length 198;
Best Local Similarity 41.7%; Pred. No. 7,4e-27;
Matches 86; Conservative 25; Mismatches 40; Indels 55; Gaps 6;

```

```

QY 13 KLLIGDSGVKSCLLIRPADDTYESYISTIGVDFKRTIELDGKTIKLOI----- 64
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 KVLIGDSGVKSSLLIRFTDNKFEVEEYIPTIGVDYTKTVEDGKTVKLIQIMDTAGGER 60
QY 65 -----ESFNVKOMLOEIDRYA--SENVNKLIVGNKCDLT- 97
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 FRALRPAYYRGAGCFLLVYDITSRDSFENVKMKLEILRHADKDENVPIVYGNKCDLED 120
QY 98 -----TKKYVDYTTAKEFADSLG-IPLETSAKNAVNEQSFMTAAEIKRRMGPG 147
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 DEDLELEGQKRVYSEEGEALAKELGALPFMETSAKTNTVNEAEFELAREILIKVY--- 177
QY 148 ATAGAERSNVKIQSTPVKQSGGCC 173
DB 178 -----SEVNVNLDQ-PAKKKRSKCC 196

```

```

RESULT 12
US-09-945-173-5
: Sequence 5, Application US/09945173
: Patent No. US20020127568A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
: FILE REFERENCE: 38155-20035.00
: CURRENT APPLICATION NUMBER: US/09/945,173
: PRIOR FILING DATE: 2001-08-31
: PRIOR APPLICATION NUMBER: US 60/229,293
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0.
: SEQ ID NO 5
: LENGTH: 198
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus amino acid sequence
US-09-945-173-5

```

```

Query Match      40.9%; Score 362.5; DB 10; Length 198;
Best Local Similarity 41.7%; Pred. No. 7,4e-27;
Matches 86; Conservative 25; Mismatches 40; Indels 55; Gaps 6;
QY 13 KLLIGDSGVKSCLLIRPADDTYESYISTIGVDFKRTIELDGKTIKLOI----- 64
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 KVLIGDSGVKSSLLIRFTDNKFEVEEYIPTIGVDYTKTVEDGKTVKLIQIMDTAGGER 60
QY 65 -----ESFNVKOMLOEIDRYA--SENVNKLIVGNKCDLT- 97
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 FRALRPAYYRGAGCFLLVYDITSRDSFENVKMKLEILRHADKDENVPIVYGNKCDLED 120

```

```

QY 98 -----TKKYVDYTTAKEFADSLG-IPLETSAKNAVNEQSFMTAAEIKRRMGPG 147
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 DEDLELEGQKRVYSEEGEALAKELGALPFMETSAKTNTVNEAEFELAREILIKVY--- 177
QY 148 ATAGAERSNVKIQSTPVKQSGGCC 173
DB 178 -----SEVNVNLDQ-PAKKKRSKCC 196

```

```

RESULT 13
US-09-972-529-4
: Sequence 4, Application US/09972529
: Patent No. US20020150916A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: TITLE OF INVENTION: 47316, A NOVEL HUMAN G-PROTEIN AND USES
: FILE REFERENCE: 38155-20041.00
: CURRENT APPLICATION NUMBER: US/09/972,529
: PRIOR FILING DATE: 2002-04-08
: PRIOR APPLICATION NUMBER: US 60/237,716
: PRIOR FILING DATE: 2000-10-05
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 198
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus amino acid sequence
US-09-972-529-4

```

```

Query Match      40.9%; Score 362.5; DB 10; Length 198;
Best Local Similarity 41.7%; Pred. No. 7,4e-27;
Matches 86; Conservative 25; Mismatches 40; Indels 55; Gaps 6;
QY 13 KLLIGDSGVKSCLLIRPADDTYESYISTIGVDFKRTIELDGKTIKLOI----- 64
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 KVLIGDSGVKSSLLIRFTDNKFEVEEYIPTIGVDYTKTVEDGKTVKLIQIMDTAGGER 60
QY 65 -----TKKYVDYTTAKEFADSLG-IPLETSAKNAVNEQSFMTAAEIKRRMGPG 147
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 FRALRPAYYRGAGCFLLVYDITSRDSFENVKMKLEILRHADKDENVPIVYGNKCDLED 120
QY 98 -----TKKYVDYTTAKEFADSLG-IPLETSAKNAVNEQSFMTAAEIKRRMGPG 147
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 DEDLELEGQKRVYSEEGEALAKELGALPFMETSAKTNTVNEAEFELAREILIKVY--- 177
QY 148 ATAGAERSNVKIQSTPVKQSGGCC 173
DB 178 -----SEVNVNLDQ-PAKKKRSKCC 196

```

```

RESULT 14
US-09-350-874-67
: Sequence 67, Application US/09350874
: Patent No. US20020019020A1
: GENERAL INFORMATION:
: APPLICANT: Rhodes, Kenneth
: TITLE OF INVENTION: METHODS FOR TREATING CARDIOVASCULAR DISORDERS
: FILE REFERENCE: NMI-069
: CURRENT APPLICATION NUMBER: US/09/350,874
: PRIOR FILING DATE: 1999-07-09
: EARLIER APPLICATION NUMBER: USSN 60/110,277
: EARLIER FILING DATE: 1998-11-30
: EARLIER APPLICATION NUMBER: USSN 60/110,033
: EARLIER FILING DATE: 1998-11-25
: EARLIER APPLICATION NUMBER: USSN 60/109,333
: EARLIER FILING DATE: 1998-11-20
: EARLIER APPLICATION NUMBER: USSN 09/298,731

```

EARLIER FILING DATE: 1999-04-23  
 NUMBER OF SEQ ID NOS: 72  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO 67  
 LENGTH: 212  
 TYPE: PRT  
 ORGANISM: Rattus sp.  
 US-09-350-874-67

Query Match 37.6%; Score 333; DB 10; Length 212;  
 Best Local Similarity 34.8%; Pred. No. 5, 1e-24;  
 Matches 73; Conservative 38; Mismatches 55; Indels 44; Gaps 2;

QY 8 YDLEKLLIGDSGVKSCLLRFADDTYESISTGVDFKIRTELDGKTIQI---64  
 Db 3 YVLFKYLIIIDTGVKSCLLQFTDKRQPVHDLTIGVEFGARMTIDGKQIKIQIMDT 62  
 QY 65 -----ESFNWKQWQLEIDRYASENVNKLVGKCD 95  
 Db 63 AGQESFRSITRSYKAGACALLVYDITRDTFNHLTTLWLEDAKQHSNKNVIMLIGNKSD 122  
 QY 96 LTTKKVVDYTTAKFADSLGIPLETSAKNATNVEQSFYTMALIKRMGPGATAGAEK 155  
 Db 123 LESRRREVKKKEGAEAFAREHGLIFMETSATASVGEAFINTAKEIYEKIQEGVFDINNEA 182  
 QY 156 SNVKIOSTPV-----KSGGGCC 173  
 Db 183 NGIKIGPOHAATNASHGNGGQAGGGCC 212

RESULT 15  
 US-09-817-198A-4  
 Sequence 4, Application US/09817198A  
 Patent No. US20020146758A1  
 GENERAL INFORMATION:  
 APPLICANT: YE, Jane et al.  
 TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
 TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
 FILE REFERENCE: CL001188  
 CURRENT APPLICATION NUMBER: US/09/817,198A  
 CURRENT FILING DATE: 2001-03-27  
 NUMBER OF SEQ ID NOS: 33  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 4  
 LENGTH: 212  
 TYPE: PRT  
 ORGANISM: Rattus norvegicus  
 US-09-817-198A-4

Query Match 36.5%; Score 323; DB 10; Length 212;  
 Best Local Similarity 34.2%; Pred. No. 4, 6e-23;  
 Matches 69; Conservative 36; Mismatches 61; Indels 36; Gaps 2;

QY 4 MNPEDYLFKLLIGDSGVKSCLLRFADDTYESISTGVDFKIRTELDGKTIQI 63  
 Db 1 MAKQYDYLFRLLIGDSGVKSCLLRFADDTYESISTGVDFKIRTELDGKTIQI 60  
 QY 64 I-----ESFNWKQWQLEIDRYASENVNKLVG 91  
 Db 61 IMDTAGQEHYQITKQYRAGSIFLYVDISSESYQHIMKNVSDVDEAPAGVQKILIG 120  
 QY 92 NNCDDLTTKVVVDYTTAKFADSLGIPLETSAKNATNVEQSFYTMALIKRMGPG 147  
 Db 121 NKADEQKROKROVREGQOLAKKEYGMDYFETSACTNINIKESFTRLTELVLQAKRRLDGL 180  
 QY 148 ATAGAEKSNVKIOSTPVQSG 169  
 Db 181 RYCASNELALAELEDEKTEG 202

Search completed: June 18, 2003, 15:24:31  
 Job time : 33 secs



US-09-820-003b-2 (1-173) x US-09-820-003a-1 (1-1405)

```

QY 1 MetSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20
DB 180 ATGTCACGACGATGATCCGGAATGATATTTATTCAGATTACTTCTGATGGCAGACCA 239
QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspPheThrTyrThrGluSerTyr 40
DB 240 GGGGTTGGAAAGCTTCCTGCTTCTTGTGGTTGAGATGATATACATACAGAAAGCTAC 299
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60
DB 300 ATCAGACCAATGCTGTGTGATTTCAAAATAGAACTATAGCTTAGACGGGAACAATC 359
QY 61 LysLeuGlnIleGluSerPheAsnAsnValLysGlnTrpLeuGlnGluIleAspArgTyr 80
DB 360 AAGCTTCAATATAGAGTCTCTCAATATATGTAAACAGTGGCGAGGAATATGATCGTAT 419
QY 81 AlaSerGluAsnValAsnLysLeuLeuValGlyAsnLysCysAspLeuThrThrLysLys 100
DB 420 GCCAGTGAATATGTCACAAATTTGTTGAGGAAATGTGATCTACACAAAGAAA 479
QY 101 ValValAspTyrThrThrAlaLysGluPheAlaAspSerLeuGlyIleProPheLeuGlu 120
DB 480 GTAGTAGCTACACACAGCCAGAGAAATTTGCTGATTCCTTGGAAATCCGTTTGGAA 539
QY 121 ThrSerAlaLysAsnAlaThrAsnValGluGlnSerPheMetThrMetAlaAlaGluIle 140
DB 540 ACCAGTGTAAAGATGCAACGAATGTATACAGCTTTTCATGACAGATGGCAGCTAGATT 599
QY 141 LysLysArgMetGlyProGlyAlaThrAlaGlyLysGluLysSerAsnValLysIle 160
DB 600 AAAAGCCCAATGGGTCGCGAGCAACACCTGCTGCTGAGAAATCCAAATGTTAAAAAT 659
QY 161 GlnSerThrProValLysGlnSerGlyLysLysCys 173
DB 660 CAGAGCACTCCAGTCACAGCTCAGGTGAGGTCTGCTC 698

```

# RESULT 2

US-09-967-736-4

; Sequence 4, Application US/09967736

; Patent No. US2002010340A1

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

Lal, Preeti

Corley, Neil C.

Shah, Purvi

TITLE OF INVENTION: RAB PROTEINS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/967,736

FILING DATE: 28-Sep-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/154,602

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0367 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 925 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: LIVERUT04

CLONE: 2514506

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-967-736-4

Alignment Scores:	Pred. No.:	Length:	Matches:
Score:	1.78e-89	925	156
Percent Similarity:	76.8	80.39%	8
Best Local Similarity:	76.47%	Mismatches:	7
Query Match:	86.74%	Indels:	33
DB:	10	Gaps:	2

US-09-820-003b-2 (1-173) x US-09-967-736-4 (1-925)

QY 2 SerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGly 21

DB 60 GCCGCGCATGAACCCCGAATATGACTTACCTGTTAAGCTGCTTTGATGGCGACTCAGGC 119

QY 22 ValGlyLysSerCysLeuLeuLeuArgPheAlaAspPheThrTyrThrGluSerTyrIle 41

DB 120 GTGGGCAAGTATGCTGCTCTCTCTGCGGTTTCTGATGACAGTACACAGAGCTTACATC 179

QY 42 SerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61

DB 180 AGCACCATCGGGGTGAGTCTCAAGATCCGAAACATCGAGCTGATGGCAAAATATATCAA 239

QY 62 Length: 64

DB 240 CTTCAGATCTGGACACACAGCGGGCCAGAGACGTTCCGACCATCTTCAGTACTTAC 299

QY 65 -----GluSerPheAsnAsn 69

DB 300 CGGGGGGCTCATGCGATCATGCTGTGATGACGCTACCTGACAGGAATCTTACGCCAAC 359

QY 70 ValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89

DB 360 GTGAAGCAGTGGCTGACAGAGATGACCGCTATGCCAGCGAAGCTCAATTAAGCTCTG 419

QY 90 ValGlyAsnLysCysAspLeuThrThrLysValValAlaAspTyrThrThrAlaLysGlu 109

DB 420 GTGGGCAACAGAGCGACCTCACACAGAAAGGTGGGACACACACAGCCAGAGAGAG 479

QY 110 PheAlaAspSerLeuGlyIleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnVal 129

DB 480 TTTCAGACTCTCGGGGATCCCTCTTGTGAGAGAGAGCGCAAGAAATGCCACCAATGTC 539

QY 130 GluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr 149

DB 540 GAGCAGCGCTTCATGACCATGCTGCTGAATCAAAAAGCGAGTGGGCGCTGAGCAGCC 599

QY 150 AlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 169

DB 600 TCTGGGGGCG---GAGCGGGCCCAATCTCAAGATGACAGACACCCCTGTAAAGCCGGCTGC 656

QY 170 GlyLysCysCys 173

DB 657 GGTGCTGTTC 668

RESULT 3

US-10-102-806-48

; Sequence 48, Application US/10102806

; Publication No. US20030054421A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.



```
DB 481 GGTTCATGCGCATGCTCTGCAACCAAGACGATGCG---AGCCACCTGCAGCA 537
QY 152 GYALAGLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 171
DB 538 GGAATCAAGCCACCAAGGTCGATCCAGAGCAACCTGTTAACGACA---TCAGGC 594
QY 172 CysCys 173
DB 595 TCCTGC 600

RESULT 5
US-09-770-445-529
; Sequence 529, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurdan, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 202305 (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 529
; LENGTH: 881
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-529

Alignment Scores:
Pred. No.: 2,71e-70 Length: 881
Score: 619.50 Matches: 125
Percent Similarity: 68.97% Conservative: 15
Best Local Similarity: 61.58% Mismatches: 30
Query Match: 69.92% Indels: 33
DB: 10 Gaps: 2

US-09-820-003B-2 (1-173) x US-09-770-445-529 (1-881)
QY 3 SerMetAsnProGlnIleuArgPheAlaAspThrTyrThcGlnSerTyrIleSer 22
DB 57 ACCATGAAATCCGAGTACACATCTTTTCAAGCTCTGCTATCGGGGATTCGCGCTA 116
QY 23 GlyLysSerCysLeuLeuLeuArgPheAlaAspThrTyrThcGlnSerTyrIleSer 42
DB 117 GCGAAGCTCTGCTCTTCTTTGAGATTCCTGATGATCTTAAGTGAAGATTCATAC 176
QY 43 ThrIleGlyValAspPheLysIleArgThrIleGlnLeuAspGlyLysThrIleLysLeu 62
DB 177 ACTAATGAGATGCGATTTTAAATATAGGACTGTGAACAAGATGGCAAAACATTAAG 236
QY 63 GlnIle----- 64
```

```
DB 237 CAAATTTGGACACTGCTGTGTCAGAACGTTTCAGACTATTACTAGACTTACCGT 296
QY 65 -----GlnSerPheAsnVal 70
DB 297 GGGGACATGAAATTATTATTGTCTACAGATGTCACAGATGAAGAAAGCTTCATATATGTC 356
QY 71 LysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuVal 90
DB 357 AAGCAATGGTTGAGTGAATGATGCTTATGCTAGTACAAATGTCACAAACTCCGTGTT 416
QY 91 GlyAsnLysCysAspLeuThrThrLysValValAspTyrThrAlaLysGluPhe 110
DB 417 GGAACACGCTGATCTTCTGAAACACAGACCATCTCTATGAACCTCCAAAGGCTTTT 476
QY 111 AlaAspSerLeuGlyIleProPheLeuGluTrpSerAlaLysAsnAlaThrAsnValGlu 130
DB 477 GCGATGAATAATCGGGATCTCTTTATGAGACTAGTGAAGAAAGTCTACAAACGTAGAA 536
QY 131 GlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThrAla 150
DB 537 CAGGCTTCATGGCAATGTCGTGATCCATCAAGAGAGATGCTAGCCACACAGCTGGC 596
QY 151 GlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 170
DB 597 AATAATGCAAGACCAACCGACCGTCGATCAGAGACAGACCTGTGACACAGAG---AAC 653
QY 171 GlyCysCys 173
DB 654 GCGTCTGC 662

RESULT 6
US-09-820-003A-3
; Sequence 3, Application US/09820003A
; Patent No. US20020142382A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001196
; CURRENT APPLICATION NUMBER: US/09/820,003A
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 46050
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(46050)
; OTHER INFORMATION: n = A,T,C or G
US-09-820-003A-3

Alignment Scores:
Pred. No.: 1,03e-50 Length: 46050
Score: 487.50 Matches: 113
Percent Similarity: 51.07% Conservative: 6
Best Local Similarity: 48.50% Mismatches: 16
Query Match: 55.02% Indels: 99
DB: 10 Gaps: 2

US-09-820-003B-2 (1-173) x US-09-820-003A-3 (1-46050)
QY 24 LysSerCysLeuLeuLeuArgPheAlaAspThrTyrThcGlnSerTyrIleSerThr 43
DB 42782 AGCTCAGCAACGATGTTCTGTTATTCGACACTATATATATATATATATACAGTC 42841
QY 44 IleGlyValAspPheLysIleArgThrIleGlnLeuAspGlyLysThrIleLysLeuGln 63
DB 42842 TTTCAAATGCTCTTT----- 42856
```



```
OY 64 Ilegiserpheasnsanvalysglntrpleuglnluileasparglyralaserglu 83
DB 42857 TAGGAGTCCTTCATATATGTAAACAGTGCCTCAGAAATAGATTCCTATGCCAGTGA 42916
OY 84 AsnValAsnValysleuValAlGlyAsnValCysAspLeuThrThrLysValAlAsp 103
DB 42917 AATGTCAACAAATGTGTAGGAAACAAATGATCTACACAAAGAAAGTAGTAGAC 42976
OY 104 ThrLysLys-----
DB 42977 TACAAACAGGAA--GGTATGTTTAAAGTTTATTTTCACTGATTTGAAGGTGTGA 43035
OY 108 -----
DB 43036 ATTATGTATGCTTCGACGTAAAGTAGCCACAGCCTTTAAATAATGTGCACTAG 43095
OY 108 -----
DB 43096 AATACTGTGACAGTGAATTTGTGTAGCATCTGTTGGATCCAACTAGTTCCCTC 43155
OY 108 -----
DB 43156 ACGCTCATATGATGATGTAGAAATGCAATAGAAATTAAGATTTTTCAGTCTT 43215
OY 109 -----
DB 43216 AATTCGCTCATATATTCCTTCTTGAATTTGCTGATTCCTTGAATTCCTTTTGGAA 43275
OY 121 ThrSerAlaLysAsnAlaThrAsnValAlGlnSerPheMetThrMetAlaAlaGlu 140
DB 43276 ACCAGTGTAAAGATCAACAGAAATGTAGAAACATCTTATGACAGATGAGCGTGAAT 43335
OY 141 LysLysArgMetGlyProGlyAlaThrAlaGlyAlaGluLysSerAsnValLysIle 160
DB 43336 AAAAAGCAATGGGTCCCGAGCAACAGCTGGTGGTGGAGAAAGTCCATGTTAAAT 43395
OY 161 GlnSerThrProValLysGlnSerGlyGlyCysCys 173
DB 43396 CAGACACTCCACGTCAAGACAGTGGAGGTGCTGC 43434

RESULT 7
US-09-820-003A-36
: Sequence 36, Application US/09820003A
: Patent No. US20020142382A1
: GENERAL INFORMATION:
: APPLICANT: MERKIOV, Gennady et al.
: TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
: FILE REFERENCE: C1001196
: CURRENT APPLICATION NUMBER: US/09/820,003A
: NUMBER OF SEQ ID NOS: 39
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 36
: LENGTH: 601
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-820-003A-36

Alignment Scores:
Pred. No.: 1,12e-52 Length: 601
Score: 481.00 Matches: 107
Percent Similarity: 56.02% Conservative: 0
Best Local Similarity: 56.02% Mismatches: 1
Query Match: 54.29% Indels: 84
DB: 10 Gaps: 1

US-09-820-003B-2 (1-173) x US-09-820-003A-36 (1-601)
OY 66 SerPheAsnValAlGlyAlaThrLysGlnLysAlaAspArgTyrAlaSerGluAsnVal 85
DB 3 TCCCTCATATATGTTTAAACAGTGCCTCAGAAATAGATCTGATGCCAGTGAATATGTC 62
```

```
OY 86 AsnValysleuValAlGlyAsnValCysAspLeuThrThrLysValAlAspTyrThr 105
DB 63 AACAAATTTGTGTGGACAAATGATCTGACCAACAAAGAAATGATAGACATACACA 122
OY 106 ThrLysLys-----
DB 123 ACACGAA--GGTATGTTTAAAGTTTATTTTCACTGATTTGAAGGTGTGAATATG 161
OY 108 -----
DB 162 TATGGTTCGACGTAAACAGTAGCCACAGCCTTTAAATAATGTGCATAGAAACT 241
OY 108 -----
DB 242 GTGACAGTCAATTTGTGTAGCATCTGTTGATCCAAATGAACTAGTTCACGCTY 301
OY 108 -----
DB 302 CATATGATGTAGAAATGCAATAGTAAATAATTTTTCAGTGTAAATGT 361
OY 109 -----
DB 362 GCCTCATATATCTCTTGAATTTGCTGATTCCTTGAATTCCTTTTGGAAACAGT 421
OY 123 AlaLysAsnAlaThrAsnValAlGlnSerPheMetThrMetAlaAlaGluLysLys 142
DB 422 GCTAAGATGCAACCAATGTAGAAACATCTTTCATGACATGCGACCTGATTTAAAG 481
OY 143 ArgMetGlyProGlyAlaThrAlaGlyAlaGluLysSerAsnValLysIleGlnSer 162
DB 482 CGAATGGGTCCCGAGCAACAGCTGGGTGCTGAGAAAGTCCAAATTTAAATTCAGAGC 541
OY 163 ThrProValLysGlnSerGlyGlyCysCys 173
DB 542 ACTCAGTCAAGCATCAGTGAAGGTGCTGC 574

RESULT 8
US-09-938-842A-836
: Sequence 836, Application US/09938842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kreps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPI300-3
: CURRENT APPLICATION NUMBER: US/09/938,842A
: NUMBER OF SEQ ID NOS: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227,866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264,647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300,111
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 5379
: SEQ ID NO 836
: LENGTH: 651
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-938-842A-836

Alignment Scores:
Pred. No.: 7,41e-46 Length: 651
Score: 428.50 Matches: 92
Percent Similarity: 58.83% Conservative: 31
Best Local Similarity: 44.02% Mismatches: 39
Query Match: 48.36% Indels: 47
DB: 9 Gaps: 5

US-09-820-003B-2 (1-173) x US-09-938-842A-836 (1-651)
```

```

OY 7 GluYrAspTyrLeuPheLysLeuLeuIleGlyAspSerGlyValGlyLysSerCys 26
DB 31 GATACGATTACCTATTAACCTCTGTCGATCGGAGACGGGTGTGTAAGATTC 90
OY 27 LeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyrIleSerThrIleGlyVal 46
DB 91 CTTCCTACGTTTCTTCAGATGCTGTTTACCACAGTTTCATTACAATTTGGGAT 150
OY 47 AspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGlnIle 64
DB 151 GATTTTAAGATACGGACATTTGAGCTCGATGGGAGAGATTAGCTGCATAATCTGGAT 210
OY 64 ----- 64
DB 211 ACTGCCGACAGAGCGGTCCGCAATCACAACCTGCTACTACCGTGGAGCCATGGGG 270
OY 65 -----GluSerPheAsnAsnValLysGlnTyrPleu 74
DB 271 AATTTGCTGTGTGANGTGTGACGTATGATGATCATCTTTCACACATCAGGAATTGGATC 330
OY 75 GlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuLeuValGlyAsnLysCys 94
DB 331 CGTACATTTGACGACATGCTCTCTGATGTCACACAGATTCTGATTTGGGAACAAGCA 390
OY 95 AspLeu---ThrThrLysLysValAlaAspTyrThrThrAlaLysGluPheAlaAspSer 113
DB 391 GATATGATGANAAGAAAGACCTGTGCCAATAATCTAAGGCCAAGCTCTGCAGATGAA 450
OY 114 LeuGlyIleProPheLysGluThrSerAlaLysAsnAlaThrAsnValGluGlnSerPhe 133
DB 451 TATGAAATGAAGCTTTTTCGAGCTAGTGTGCCAAGACTAACTTAACCTTGAAGAAAGTTTC 510
OY 134 MetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyValAlaThrAlaGlyAla 153
DB 511 TTCTGTATTCATAAGACATTAAGCAACACTTGCAGAT-----ACCGATGACAGAGCT 564
OY 154 GluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGlyGly----- 171
DB 565 GAGCGCGCAAAACATCAAAATCAACCAATCC-----GACCAAGGTGCGGGAAACATCT 615
OY 172 -----CysCys 173
DB 616 CAGCTACTCAGAAATCAGCATGTTGC 642

```

RESULT 9  
 US-09-794-257-9  
 ; Sequence 9, Application US/09794257  
 ; Patent No. US20020009804A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meyers, Rachel  
 ; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1e1  
 ; FILE REFERENCE: 35800/209285  
 ; CURRENT APPLICATION NUMBER: US/09/794,257  
 ; PRIOR FILING DATE: 2001-02-27  
 ; PRIOR APPLICATION NUMBER: 60/185,606  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9  
 ; LENGTH: 624  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 US-09-794-257-9

Alignment Scores:  
 Pred. No.: 5.16e-44 Length: 624  
 Score: 414.00 Matches: 87  
 Percent Similarity: 60.50% Conservative: 34  
 Best Local Similarity: 43.50% Mismatches: 41  
 Query Match: 46.73% Indels: 39  
 DB: 10 Gaps: 3

```

US-09-820-003b-2 (1-173) x US-09-794-257-9 (1-624)
OY 4 MetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGlyValGly 23
DB 1 ATGGCGACAGCGTATGATTAATCTTCAAGCTCTGCTGATGGGAGCTGGGGGTAGGC 60
OY 24 LysSerCysAlaLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyrIleSerThr 43
DB 61 AAGACCTGCTCTGTTCCGCTTCACAGAGAGCGGCTTAACACACACTTATCTCCACC 120
OY 44 IleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGln 63
DB 121 ATCGGAATGATTTAAATTAAGACGATAGACGAAAGAAATTAAGCTTCAG 180
OY 64 Ile----- 64
DB 181 ATATGGACACAGCGGGGTGAGAAAGATTCCGAAACATCAGACAGCTTACTACAGAGA 240
OY 65 -----GluSerPheAsnAsnValLys 71
DB 241 GCCATGGCGATTATGCTGTATGATCATCACAATGAAATTCCTTGACATATTAA 300
OY 72 GlnTyrPheGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuValGly 91
DB 301 AATTTGATCAGAAATGTAAGACATGCTCTTCCGATGTCGAAAGAAATGATCCTGGGT 360
OY 92 AsnLysCysAspLeuThrThrLysLysValAlaAspTyrThrThrAlaLysGluPheAla 111
DB 361 AACAAATGATATGATGATGACAAAGACAGTGTCAAAAGAAAGGGGAGGAGCTGACA 420
OY 112 AspSerIleGlyIleProPheLysGluThrSerAlaLysAsnAlaThrAsnValGluGln 131
DB 421 ATTGACTATGATTAATATTTCTTGAGACAAACGCAAAATCCAGTCAAAATGTAGAAAG 480
OY 132 SerPheMetThrMetAlaAlaGluIle-----LysLysArgMetGlyProGly 147
DB 481 GCATTTTTCACCTTGCACGAGATATTAATGACAAAACCTCAACAGAAATGATGACAGC 540
OY 148 AlaThrAlaGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGln 167
DB 541 AATTCAGCAGAGCAGCT-----GGACCGATGTAATTAACGAAAA-CGATTCAGAA 593

```

RESULT 10  
 US-09-794-257-7  
 ; Sequence 7, Application US/09794257  
 ; Patent No. US20020009804A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meyers, Rachel  
 ; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1e1  
 ; FILE REFERENCE: 35800/209285  
 ; CURRENT APPLICATION NUMBER: US/09/794,257  
 ; PRIOR FILING DATE: 2001-02-27  
 ; PRIOR APPLICATION NUMBER: 60/185,606  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 1161  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 ; NAME/KEY: CDS  
 ; LOCATION: (18)...(641)  
 US-09-794-257-7

Alignment Scores:  
 Pred. No.: 1.3e-43 Length: 1161  
 Score: 414.00 Matches: 87  
 Percent Similarity: 60.50% Conservative: 34  
 Best Local Similarity: 43.50% Mismatches: 41  
 Query Match: 46.73% Indels: 39

DB: 10 Gaps: 3

US-09-820-003b-2 (1-173) x US-09-794-257-7 (1-1161)

OY 4 MetasnpProglutTyraSPYrLeuphelaLysLeuleuilegIyaspserGIyalaGly 23  
|||  
DB 18 ATGGGAGAGACGATATATCTCTCAACCTCTCTGATGCGGAGCTGGGGGTAGGC 77  
OY 24 LysSerCysLeuleuLeuArpPhelaAspAPThrTYrThrgIuseryTylieserthr 43  
|||  
DB 78 AAGACCTGCTCTCTGCTTCAGAGAGAGCGCTTCACACACACCTCATCTCCACC 137  
OY 44 IllegIyalaSpphelaSylearThrIllegIuleuaspGlyLysThrIlleYleuGln 63  
|||  
DB 138 ATCGGAATGATTTTAAATATGAACGATGAACTGATGAAAGAAATTAAGCTTCAG 197  
OY 64 Ille----- 64  
DB 198 ATATGGACACACAGCGGCTCAGAAAGATTCGAAACATCAGACAGCGTACTACAGAGA 257  
OY 65 -----GluSerPheaspnsValLys 71  
DB 258 GCCATGGCATTATGCTGCTATGACATCAAAATGAAATATCTTGTGCAATATTAA 317  
OY 72 GlnTrpLeuGlnGluIleaspArgTYrAlaSerGluaspValaspLysLeuleuValGly 91  
|||  
DB 318 AATTTGGATCGAAACATTCAGAGACATGCTCTCCGATGTCGAAAGAAATGATCTCTGGCT 377  
OY 92 AsnLysCysaspLeuThrThrLysLysValaspTYrThrThralaLysGluPheala 111  
|||  
DB 378 AACCAATGTATGTATGATGACAAAGACAGTCTCAAAAGAAAGGGGAGAGAGCTAGCA 437  
OY 112 AspSerLeuGlnIleProPheLeuGlnThrSerAlaLysAsnAlaThrasnValGlnGln 131  
|||  
DB 438 ATTGACTATGAGATTAATCTGTGAGACAAAGCGCAAAATCCATGCAAAATGAGAGAG 497  
OY 132 SerPheMetThrMetAlaIle-----LysLysArgMetGlyProGly 147  
|||  
DB 498 GCATTTTACTACTGTCAGAGATATATGACAAACCTCAACAAATAATGATGACAGC 557  
OY 148 AlaThrAlaGlyLysLysSerAsnValLysIlleGlnSerThrProValLysGln 167  
|||  
DB 558 AATTCAGCAGACAGAGT-----GACACAGTGAATAATACAGAAA-CCGATCAAGAA 610

RESULT 11

US-09-834-975-879 Application US/09834975

Sequence 879 Patient No. US20020110815A1

GENERAL INFORMATION:

APPLICANT: Lillie, James

APPLICANT: Brown, Jeffrey

APPLICANT: Bolt, Andrew

APPLICANT: Van Huffel, Christophe

TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS

TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

FILE REFERENCE: MRI-016B

CURRENT APPLICATION NUMBER: US/09/834,975

PRIOR FILING DATE: 2001-04-13

PRIOR APPLICATION NUMBER: 60/197,538

PRIOR FILING DATE: 2000-04-14

NUMBER OF SEQ ID NOS: 1046

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 879

LENGTH: 2497

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

LOCATION: (1) (2497)

OTHER INFORMATION: n - A,T,C or G

US-09-834-975-879

Alignment Scores:

Pred. No.: 4,06e-43 Length: 2497

Score: 87 Matches: 414.00

Percent Similarity: 60.50% Conservative: 34

Best Local Similarity: 43.50% Mismatches: 41

Query Match: 46.73% Indels: 39

DB: 10 Gaps: 3

US-09-820-003b-2 (1-173) x US-09-834-975-879 (1-2497)

OY 4 MetasnpProglutTyraSPYrLeuphelaLysLeuleuilegIyaspserGIyalaGly 23  
|||  
DB 92 ATGGCAGAGACGATATATCTCTCAACCTCTCTGATGCGGAGCTGGGGGTAGGC 151  
OY 24 LysSerCysLeuleuLeuArpPhelaAspAPThrTYrThrgIuseryTylieserthr 43  
|||  
DB 152 AAGACCTGCTCTCTGCTTCAGAGAGAGCGCTTCACACACACCTCATCTCCACC 211  
OY 44 IllegIyalaSpphelaSylearThrIllegIuleuaspGlyLysThrIlleYleuGln 63  
|||  
DB 212 ATCGGAATGATTTTAAATATGAACGATGAACTGATGAAAGAAATTAAGCTTCAG 271  
OY 64 Ille----- 64  
DB 272 ATATGGACACACAGCGGCTCAGAAAGATTCGAAACATCAGACAGCGTACTACAGAGA 331  
OY 65 -----GluSerPheaspnsValLys 71  
DB 332 GCCATGGCATTATGCTGCTATGACATCAAAATGAAATATCTTGTGCAATATTAA 391  
OY 72 GlnTrpLeuGlnGluIleaspArgTYrAlaSerGluaspValaspLysLeuleuValGly 91  
|||  
DB 392 AATTTGGATCGAAACATTCAGAGACATGCTCTCCGATGTCGAAAGAAATGATCTCTGGCT 451  
OY 92 AsnLysCysaspLeuThrThrLysLysValaspTYrThrThralaLysGluPheala 111  
|||  
DB 432 AACCAATGTATGTATGATGACAAAGACAGTCTCAAAAGAAAGGGGAGAGAGCTAGCA 511  
OY 112 AspSerLeuGlnIleProPheLeuGlnThrSerAlaLysAsnAlaThrasnValGlnGln 131  
|||  
DB 512 ATTGACTATGAGATTAATCTGTGAGACAAAGCGCAAAATCCATGCAAAATGAGAGAG 571  
OY 132 SerPheMetThrMetAlaIle-----LysLysArgMetGlyProGly 147  
|||  
DB 572 GCATTTTACTACTGTCAGAGATATATGACAAACCTCAACAAATAATGATGACAGC 631  
OY 148 AlaThrAlaGlyLysLysSerAsnValLysIlleGlnSerThrProValLysGln 167  
|||  
DB 632 AATTCAGCAGACAGAGT-----GACACAGTGAATAATACAGAAA-CCGATCAAGAA 684

RESULT 12

US-09-834-975-885 Application US/09834975

Sequence 885 Patient No. US20020110815A1

GENERAL INFORMATION:

APPLICANT: Lillie, James

APPLICANT: Brown, Jeffrey

APPLICANT: Bolt, Andrew

APPLICANT: Van Huffel, Christophe

TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS

TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

FILE REFERENCE: MRI-016B

CURRENT APPLICATION NUMBER: US/09/834,975

PRIOR FILING DATE: 2001-04-13

PRIOR APPLICATION NUMBER: 60/197,538

PRIOR FILING DATE: 2000-04-14

NUMBER OF SEQ ID NOS: 1046

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 885

LENGTH: 2497

TYPE: DNA

ORGANISM: Homo sapiens

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2497)
; OTHER INFORMATION: n = A,T,C or G
US-09-834-975-885

```

**Alignment Scores:**

Pred. No.:	4,06	43	Length:	249
Score:	414	00	Matches:	87
Percent Similarity:	60	508	Conservative:	34
Best Local Similarity:	43	508	Mismatches:	41
Query Match:	46	73%	Indels:	39
DB:	10		Gaps:	3

US-09-820-003B-2 (1-173) x US-09-834-975-885 (1-2497)

QY 4 MetLsnProGluTyrAspTyrLeuPheLysLeuLeuLeuIleGlyAspSerGlyValGly 23  
Db 92 ATGGCGAAGACGTACGATTATCTTCACAGCTCCTGCTGATCGGCGCACTCGGGGGTAGGC 151

QY 24 LysSerCysLeuLeuLeuArgPheAlaAspSPTHTYRTHGluSerThr 43  
 |||:::||||| |||:::||||| :: |||:::||||| |||  
 Db 152 AAGACTGCTCTGTTCCGGTTTCAAGAGACGCTTCAACACCACTTCATCTCACCC 211

QY 44 IlegiValaspPhelysIleargThrIlegiIuleuaspGlyLysThrIleLysleucin 63  
|||||:::|||||  
Db 212 ATCGAATTGATTTTAAAAATTGAACGATAGAACTACATGCAAGAAGAAATTTAGCTTCAG 271

Oy 64 1le----- 64  
 ||  
 Db 272 ATATGGACACAGCGGTACGAAAGATTCCGAAACAATCAGCAGCAGCGTACTACAGAGGA 331

QY 65 -----GlutserpheasnasValys 71  
:::|||||:::|||||  
Db 332. GCCATGGGCAATTATGCTGTCATGACATCAAAATGAAAAAACCTTGACATATTAA 391

```

Oy      72 GlnTTPLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuLeuValGly 91
      . |||::: ||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db     392 AATTGGATCAAGAACATTGAAGAGACATGCCCTTCGCATGTGCAGAAAGATGATCCTGGGT 455

```

QY 92 AsnLysCysaspLeuThrThrLysLysValValAspTyrThrAlaLysGluPheAla 111  
|||||  
Db 452 AACCAATGTGATATGATGACAAAAGACAACTGTCAAAAGAAAGAGGGAGAGACTTACA 511

Oy      112 AsperleuglylIeProphelEugluThrserAlalysAsnaLaThrAsNaValGluIn 131  
         ||||| ||||||| : : : : : ||||||| :  
Db      512 ATTGACTATGGGATTAAATCTTGGAGACACAACCGCAAAATCCAGTCGCAAATGTAGNAGAG 571

```

QY      132 SerPheMetThrMetAlaAlaGluLe-----LysLysArgMetGlyProGly 147
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      572 GCATTTTTTCACCTTGCACGCGATATTAATGACAAAACTCAACAGAAAAATGAATGACACG 631

```

QY	148	AlaThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnInsertHphProValLysGln	167
Db	632	AATTCAGCAGCAGGAGCAGT-----GCACCAGTGAAATAATACAGAAA-CCGATCAAGAA	644

### RESULT 13

Sequence 894, Application US/09834975  
Patent No. US20020110815A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Brown, Jeffrey  
APPLICANT: Bolt, Andrew  
APPLICANT: Van Hufel, Christophe  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS  
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
FILE REFERENCE: MRI-016B  
CURRENT APPLICATION NUMBER: US/09/834,975  
CURRENT FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/197,538  
PRIOR FILING DATE: 2000-04-14

```

; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0.0

```

```

; ORGANISM: Homo sapiens
; FEATURE:
;

```

```

; LOCATION: (1)...(2497)
; OTHER INFORMATION: n - A,T,C or G
US-09-834-975-894

```

Alignment Scores:	
Pred. No.:	4.06e-43
Score:	414.00
Length:	2497
Matches:	87

Best Local Similarity:	43.50%	Mismatches:	41
Query Match:	46.73%	Indels:	39
DB:	10	Gaps:	3

4 MetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAsnSerGlyValGly 23

Db	92	ATGGCGAAGACGTACAGTATTACTCTTCAAGCTCCTGCGATCGGCGACTCGGGGGTAGGC	15
Oy	24	LysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyrIleSerThr	43

Db	152	AAGACCTGCTCCCTGTTCCGCTTCTAGAGAGAGCCCTCAACACACACCTTCATCTCCACC	21
		.....	
Qy	44	IlleglValAspPheIysIlleArgThrIlleglIleuAspGlyIysThrIleIysLeuGln	63

Db 212 ATCGAATTGATTTTAAAAATTAGAACATAGACTGGAAGAATAATTAGCTTCAG 27

Qy 64 Ile..... 64

Db	272	ATATGGGACACAGCGGTCAGGAAAGATTCCGACAAATCAGCAGCGTACTACAGAGGA	33
Oy	65	-----GluSerPheAsnAsnValLys	71

```

Db      332  GCCATGGGCATTATGCTGTTATGACATCACAATGAAAAATCCTTGACAATATTAA 39
QY      72  GlnTrpLeuGlnIuIeAspArqTYrAlaSerGluhsnValIsnLysLeuValGly 91

```

Db 392 AATTGGATCAGAACATTGAGAGCATGCCCTCTTCCGATGTCCAAAGATGATTCCTGGGT 45

Db 452 AACAAATGTCATATGATGACAAAGCACAGTGTCAAAGAAAGAGGGGAGAAAGCTAGCA 511

```

Db      512  ATTCGCTATGGGATTAAATTCCTTGGAGACAACGCCAAAATCCAGTGCAAATGTGAGAGAG 57
      |||||
Ov      132  SerpHeMethrMeta|aa|ag|u|le-----[vslvsaYromatG|vbrG|lv 14

```

Db

Oy 148 AlcthrAlaGlyGlyAlaGlyIuLySeranValIySlcIuSerThrProValIySgIn 167  
:::||||| ||||||| ||:::||||::

Db 632 AATTCACGACGACGAGT-----GGACCAGTGAATAATACAGAAA-CCGATCAAGAA 684

RESULT 14  
US-09-834-975-896  
Sequence 896, Application US/09834975  
Patent No. US20020110815a1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Brown, Jeffrey  
APPLICANT: Bolt, Andrew  
APPLICANT: Van Huffel, Christophe  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS  
FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY

TITLE OF INVENTION: OF HUMAN CANCERS  
 FILE REFERENCE: MRI-016B  
 CURRENT APPLICATION NUMBER: US/09/834,975  
 CURRENT FILING DATE: 2001-04-13  
 PRIOR APPLICATION NUMBER: 60/197,538  
 PRIOR FILING DATE: 2000-04-14  
 NUMBER OF SEQ ID NOS: 1046  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 896  
 LENGTH: 2497  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc-feature  
 LOCATION: (1)...(2497)  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-834-975-896

## Alignment Scores:

Pred. No.:	4,06e-43	Length:	2497
Score:	414.00	Matches:	87
Percent Similarity:	60.508	Conservative:	34
Best Local Similarity:	43.508	Mismatches:	41
Query Match:	46.738	Indels:	39
DB:	10	Gaps:	3

US-09-820-003b-2 (1-173) x US-09-834-975-896 (1-2497)

QY 4 Metlanpocglutyrasptyrleuphelyleuileuilegiaspserglyvalgly 23  
 DB 92 ATGGCGAAGACGATACATATCTCTCAACCTCTCGATGCGGCGATCGGGGATAGC 151  
 QY 24 IysSerCysleuLeuLeuAtrPheAlaAspAspThrTyrrhrGuserrThrlleserThr 43  
 DB 152 AAGACCTCCCTCGCTCGCTCGCTCGAGAGAGCGCTTCAACACACACCTCATCTCCACC 211  
 QY 44 IleglValAspPheLysIleatgThrIleglLeuAspGlyLysThrIleLysLeuGln 63  
 DB 212 ATCGGATGATTTTAAATTAGAACATAGAACTAGATGAGAAAGAAATTAAGCTCAG 271  
 QY 64 Ile----- 64  
 DB 272 ATATGGACACACGCGGCTCAGAAAGATTCGAAACATCAGACAGCGTACTACAGAGA 331  
 QY 65 -----GluSerPheAsnAsnValLys 71  
 DB 332 GCCATGGCATATATGCTGTATGACATCATCAAAATGAAATCTTGTGCAATATTA 391  
 QY 72 GlnTrpLeuGlnGlnIleAspArgTyrrAlaSerGluAsnValAsnLysLeuValGly 91  
 DB 392 AATTGATATGAGAAACATTGAGAGACATGCTCTCCGATGTGCAAAATGATCTGGGT 451  
 QY 92 AsnLysCysAspLeuThrThrLysValValAspTyrrThrAlaLysGluPheAla 111  
 DB 452 AACAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511  
 QY 112 AspSerLeuGlyLleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnValGln 131  
 DB 512 ATTGACTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 571  
 QY 132 SerPheMetThrMetAlaAlaGluIle-----LysLysArgMetGlyProGly 147  
 DB 572 GCATTTTACCTTCCAGACATATATATATATATATATATATATATATATATATATAT 631  
 QY 148 AlaThrAlaGlyLysAlaGlyLysSerAsnValLysIleGlnSerThrProValLysGln 167  
 DB 632 AATTTCAGAGGACGACGT-----GCACCATGTAATAATACAGAAA-CGATCAAGAA 684

RESULT 15  
 US-09-925-300-631  
 ; Sequence 631, Application US/09925300  
 ; Patent No. US20020151681A1  
 ; GENERAL INFORMATION:

APPLICANT: Craig Rosen,  
 APPLICANT: Steve Ruden  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 FILE REFERENCE: PA101  
 CURRENT APPLICATION NUMBER: US/09/925,300  
 CURRENT FILING DATE: 2001-08-10  
 PRIOR APPLICATION NUMBER: PCT/US00/05988  
 PRIOR FILING DATE: 2000-03-08  
 PRIOR APPLICATION NUMBER: 60/124,270  
 PRIOR FILING DATE: 1999-03-12  
 NUMBER OF SEQ ID NOS: 1890  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 631  
 LENGTH: 1537  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc feature  
 LOCATION: (5)  
 OTHER INFORMATION: n equals a,t,g, or c  
 US-09-925-300-631

## Alignment Scores:

Pred. No.:	4,46e-42	Length:	1537
Score:	403.50	Matches:	85
Percent Similarity:	54.678	Conservative:	32
Best Local Similarity:	39.728	Mismatches:	30
Query Match:	45.548	Indels:	67
DB:	10	Gaps:	4

US-09-820-003b-2 (1-173) x US-09-925-300-631 (1-1537)

QY 8 TyrAspTyrrleuphelyleuileuilegiaspserglyvalglyLysSerCysLeu 27  
 DB 548 TACACCTCTTTTCAACCTGCTCTGATGCGGGGATTCGGAATGGGGAACCTGCTGC 607  
 QY 28 LeuLeuAspPheAlaAspAspThrTyrrhrGuserrThrlleserThrlleglValAsp 47  
 DB 608 CTTTCGCTTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 667  
 QY 48 PheLysIleatgThrIleglLeuAspGlyLysThrIleLysLeuGlnIle----- 64  
 DB 668 TTCAAGTCAAAAGAGCTGTAATTAACAAGAAAGAAAGAAATCAAGCTACAGATGATACA 727  
 QY 64 ----- 64  
 DB 728 GCAGCGAGAGCGATTTCACACCATCACACCTCTCTACAGAGCGCAATGCTATC 787  
 QY 65 -----GluSerPheAsnAsnValLysGlnTrpLeuGln 75  
 DB 788 ATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 847  
 QY 76 GlnIleAspArgTyrrAlaSerGluAsnValAsnLysLeuValGlyAsnLysCysAsp 95  
 DB 848 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 907  
 QY 96 LeuThrThrLysLysValValAspTyrrThrAlaLysGluPheAlaAspSerLeuGly 115  
 DB 908 ATGACGACACAAAGAGTTGATCTTAAAGAAAGAGAGAACAGATTGCAAGGACATGCT 967  
 QY 116 IleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnValGlnGlnSerPheMetThr 135  
 DB 968 ATTGGTTTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1027  
 QY 136 MetAlaAlaGlnIleLysLysArgMetGlyProGlyAlaThrAlaGlyValAlaGlyLys 155  
 DB 1028 TTACCTGCAAGATATC----- 1042  
 QY 156 SerAsnValLysIleGlnSerThrProValLysGln----- 167  
 DB 1043 -----CTTCGAAAGACCCCTGTATTAAGAGCCCAACGTAAGTAATGATATC 1090  
 QY 168 ---serglyglyly-----CysCys 173

Mon Jun 23 16:54:08 2003

us-09-820-003b-2.p2n.rnpb

Page 10

Db 1091 AGCAGTGGAGGAGGCGCTGACAGGCTGGAAGAGCAATGCTGC 1132

Search completed: June 22, 2003, 19:48:03  
Job time : 154 secs







[illegible]

US-09-399-913-66  
; Sequence 66, Application US/09399913

```

APPLICANT : Rhodes, Kenneth
APPLICANT : Betty, Maria
APPLICANT : Ling, Hui-Ping
APPLICANT : An, Wengdan
TITLE OF INVENTION : TOPICSTIM CHANNEL INTERACTORS AND USES THEREFOR
FILE REFERENCE : NMI-0700P2
CURRENT APPLICATION NUMBER : US/09/399, 913
CURRENT FILING DATE : 1999-09-21
EARLIER APPLICATION NUMBER : USSN 60/110,277
EARLIER FILING DATE : 1998-11-30
EARLIER APPLICATION NUMBER : USSN 60/110,033
EARLIER FILING DATE : 1998-11-25
EARLIER APPLICATION NUMBER : USSN 60/109,333
EARLIER FILING DATE : 1998-11-20
EARLIER APPLICATION NUMBER : USSN 09/298,731
EARLIER FILING DATE : 1999-04-23
EARLIER APPLICATION NUMBER : USSN 09/350,614
EARLIER FILING DATE : 1999-07-09
EARLIER APPLICATION NUMBER : USSN 09/350,874
NUMBER OF SEQ ID NOS : 73
SOFTWARE : PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH : 639
TYPE : DNA
ORGANISM : Rattus sp.
FEATURE :
NAME/KEY : CDS
LOCATION : (1)..(636)
OS-09-399-913-66

```

### Alignment Scores:

Pred. No.:	2.52e-36	Length:	639
Score:	333.00	Matches:	73

Percent Similarity:	52.86%	Conservative:	3
Best Local Similarity:	34.76%	Mismatches:	5
Query Match:	27/50		7-3-3

```
Query-Match: 37.58% Indels: 4
DB: 4 Gaps: 2
```

US-09-820-003B-2 (1-173) x US-09-399-913-66 (1-639)

8 TyrAspTyrLeuPheIysLeuLeuIleGlyAspSerGlyValGlyIysSerCysLeu 27

Db 7 TACGCCATCTCTTCAAGTACATCATCATCGGCGACACAGGTCGTTGGTAAATCGTCTTA 66

QY 28 LeuLeuArgPheAlaAspAspThrTyrThrGluSerTyrIleSerThrIleGlyValAsp 47  
|||||:::||| |||||:::||| |||||:::||| |||||:::||| |||||:::|||

Db 67 TTGCTACAGTTTACAGACAGAGGTTTACGCCGTGCTGACTTACACTTGCATTTGGTGTAG 126

[illegible]

RESULT 4  
US-08-888-077A-28

US 08 888 077A-28  
; Sequence 28, Application US/08888077A  
; Patent No. 6020143

APPLICANT: ST. GEORGE-HYSIOP, PETT

APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E

INVENTOR	INVENTOR'S ADDRESS	TITLE OF INVENTION:	GENETIC SEQUENCE
;	;	TITLE OF INVENTION:	TO ALZHEIMER'

NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LERNER, DAVID, LITTE  
STREET: 600 SOUTH AVENUE WEST

```

; CITY: WESTFIELD
; STATE: NJ
;

```

; COUNTRY: USA  
; ZIP: 07090-1497  
;

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

;
;  COMPUTER:  IBM PC compatible
;  OPERATING SYSTEM:  PC-DOS/MS-DOS
;

```

```

; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
;

```

APPLICATION NUMBER: US/08/888,0  
FILING DATE: 03-JUL-1997

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; ADDITIONAL NUMBER: TSC 004003 E

```

APPLICATION NUMBER: US 08/592,5  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:  
NAME: PALISI, THOMAS M  
REGISTRATION NUMBER: 35 639

REGISTRATION NUMBER: 36,029  
REFERENCE/DOCKET NUMBER: . SCHERTI  
TELECOMMUNICATION INFORMATION:

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 654-5000 .

TELEFAX: (908) 654-7866  
 INFORMATION FOR SEQ ID NO: 28:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 970 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 1..970  
 OTHER INFORMATION: /note="Y2H9"  
 US-08-888-077A-28

Alignment Scores:  
 Pred. No.: 3,92e-34 Length: 970  
 Score: 319.00 Matches: 77  
 Percent Similarity: 52.11% Conservative: 34  
 Best Local Similarity: 36.15% Mismatches: 62  
 Query Match: 36.00% Indels: 40  
 DB: Gaps: 3

US-09-820-003B-2 (1-173) x US-08-888-077A-28 (1-970)

```

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuLeuIleGlyAspSer 20
DB 54 ATGGGACCCGCGAGACGAGTACGACTCTTAAAGTGTCTTATGGAGATCT 113
QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspPThrTyrThrGluSerTyr 40
DB 114 GGTGTGGAAAGATGATCTCTGCTGATTTACTCGAATGATTTAATGTGAAAGC 173
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60
DB 174 AAGACACCATGATGGATGAGTGTTCACAAAGAACATCCAGGTTGATGAAACAAATA 233
QY 61 LysLeuGlnIle----- 64
DB 234 AAGGACAGATATGGGACACAGCAGGACAGCATATCCAGCTATATACATCAGCATAT 293
QY 65 -----GluSerPheAsn 68
DB 294 TATCGTGAGCTGTAGTGTCTTATTGGTTATGACATTGCTAAACATCTCAGATATGAA 353
QY 69 AsnValLysGlnTyrLeuGlnIleAspArgTyrAlaSerGluAsnValAsnLysIleu 88
DB 354 AATGATAGGCGATGCTGTAAGAACTGAGATCATGCTGATATGATGATGTTATCATG 413
QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108
DB 414 CTTGTGGCATATAGAGTATGATCTATCTCAGGCAAGTTCTTACAGATGAAAGCAAGA 473
QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGlnTyrSerAlaLysAsnAlaThrAsn 128
DB 474 GCTTTGCAAGAAAGATGTTGTCATTCATTGAATCTGGCCCTAGACTCTCAAAAT 533
QY 129 ValGlnGlnSerPheMetThrMetAlaIle-----LysLysArg 143
DB 534 ATAGAGCGTGTCTTTCACAGCAATTTTAAACAGAGATTTCCGCAATGTTTTCAGAGCA 593
QY 144 MetGlyProGlyAlaThrAlaGlyValGluLysSerAsnVal-----LysIle 160
DB 594 ATGTACAGACAGACGGAATGATGATCTCTCCAGCAACAAATGTGTTCTATTTCATGTT 653
QY 161 GlnSerThrProValLysGlnSerLysGlyLysCysCys 173
DB 654 CCACCAACCACTGAAACAGCAAAAGGTGACAGTCTCT 692

```

RESULT 5  
 US-09-075-454-10  
 Sequence 10, Application US/09075454  
 Patent No. 6391580  
 GENERAL INFORMATION:  
 APPLICANT: Hillman, Jennifer L.

APPLICANT: Tang, Y. Tom  
 APPLICANT: Lal, Preeti  
 APPLICANT: Guegler, Karl J.  
 APPLICANT: Corley, Neil C.  
 APPLICANT: Patterson, Chandra  
 APPLICANT: Batra, Sajeev  
 APPLICANT: Baughn, Mariah R.  
 TITLE OF INVENTION: RAS PROTEINS  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: US  
 ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: Word Perfect 6.1/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/075,454

FILING DATE: Herewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/766,551

FILING DATE: DECEMBER 12, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Cerrone, Michael C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0168-1 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 875 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: DCMCLST01

CLONE: 1528559

US-09-075-454-10

Alignment Scores:  
 Pred. No.: 1.38e-33 Length: 875  
 Score: 314.50 Matches: 61  
 Percent Similarity: 56.90% Conservative: 38  
 Best Local Similarity: 35.06% Mismatches: 42  
 Query Match: 35.50% Indels: 33  
 DB: Gaps: 2

US-09-820-003B-2 (1-173) x US-09-075-454-10 (1-875)

```

QY 5 AsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGlyValGlyLys 24
DB 63 AGTCGCACTACGACCTCAGCGGCAAGTATGCTCTGGAGACACAGGCTGGCAAA 122
QY 25 SerCysLeuLeuLeuArgPheAlaAspPThrTyr---ThrGluSerTyrIleSerThr 43
DB 123 ACATGTTTCGATTCATTCATTAAGACGCGGCTTCTCCGCACTTCATATGACCAAC 182
QY 44 IleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGln 63
DB 183 GTCCGATATGACTTCAGAGCAAGGTGTGACTGTGATGGCTGAGAGTGAAGCTGCAG 242
QY 64 Ile----- 64
DB 243 ATCTGGACACCCGCTGGGCGAGAACGGTTCGGAAGCGTACCATGCTTATTACAGAGAT 302

```

QY	65	-----	Gluserpheanaasenvallys	71
Db	303	GCCTACGCTTCCTGCTGTATGACATCACCAACAAATCTCTTCGCACAAATCAGG	365	-----
QY	72	GLTRPLeuGIuGIuIIeAspAgyrTRIAseRJuIuAsnValAsnIysLeuIeuValGly	91	-----
Db	363	GCCTGGCTACTACGAGATTCATAGAGTRTCCAGAGGAGAGCTGTGATCATCTGCTAGGC	422	-----
QY	92	AsnIysAcYasApleuThrThrIysIysValValAspTYrThrThrAlaIysGIuPheAla	111	-----
Db	423	AACACAGCGAGATATAGACACGAAAGATGTATCGCTCCGAAGACGAGACCTTGAC	482	-----
QY	112	AspSerLeuGIuIIeProPheLeuGIuThrSerAlaIysAsnAlaThrAsnValGIuGIu	131	-----
Db	483	AGGAGATAGGGGTCTCTCTCCCTCCGAGAGACAGCCCAACATCGCATGATGTGGAGTTA	542	-----
QY	132	SerPheMetThrMetAlaIaGIuIIeIysIysAsnArgMetGly	145	-----
Db	543	GCCTTCTGGCCATCGCAAGAACTGAATATACCGGCGCGG	584	-----
RESULT	6			
US-09-484-970B-142				
Sequence 142, Application US/09484970B				
Patent No. 6426186				
GENERAL INFORMATION:				
APPLICANT: Jones, Karen A.				
APPLICANT: Volkmeit, Wayne				
APPLICANT: Walker, Michael G.				
TITLE OF INVENTION: BONE REMODELING GENES				
FILE REFERENCE: PB-0014 US				
CURRENT APPLICATION NUMBER: US/09/484,970B				
CURRENT FILING DATE: 2000-01-18				
NUMBER OF SEQ ID NOS: 172				
SOFTWARE: PERL Program				
SEQ ID NO 142				
LENGTH: 2612				
TYPE: DNA				
ORGANISM: Homo sapiens				
FEATURE:				
NAME/KEY: misc.feature				
OTHER INFORMATION: Incyte ID No. 6426186 412477.1CB1				
US-09-484-970B-142				
Alignment Scores:				
Pred. No.:	7,35e-33	Lengths:	2612	
Score:	314.50	Matches:	61	
Percent Similarity:	56.90%	Conservative:	36	
Best Local Similarity:	35.00%	Mismatches:	42	
Query Match:	35.50%	Indels:	33	
Db:	4	Gaps:	2	
US-09-820-003B-2 (1-173) x US-09-484-970B-142 (1-2612)				
QY	5	AsnProGIuTYrAspTYrIleuPheIysLeuIeuIIeGIyAspSerGIyValGIyIys	24	-----
Db	82	AGTCCAGCTACGACACCTCAGCGGCAAGGTGATGCTCTGTGGAGACACAGCGCTGGCAAA	141	-----
QY	25	SerCysLeuIleuLeuIleuArgPheAlaAspAphTYr---ThsGIuSerTYrIleSerThr	43	-----
Db	142	ACATGTTCTCTGATTCACAAATTCMAAGCGGGCCCTTCGTCCGGAACCTTCATAGCAAC	201	-----
QY	44	IleGIyValAspPheIysIleIeGThrIleIeIuAspClyIysThrIleIysIeuIn	63	-----
Db	202	GTCGCGATGACATCAGAACAAAGGTGTACTGTGATGTGGCGTGAAGATGAAGCTCGAG	261	-----
QY	64	Ile-----	64	-----
Db	262	ATTCGGGACACCGCTGGGCGACGAGCGTTCGAAAGCGTACCCATGCTTATTACAGAGAT	321	-----
QY	65	-----	71	-----
Db	332	GCCTACGCTTCCTGCTGTATGACATCACCAACAAATCTCTTCGCACAAATCAGG	381	-----

```

Oy      7 GtATPLeugInIuLlAsPaTgYrAlAserCluaSnVaLaSnyLSuLeuValGly 91
Db      382 GCCTGGCTACAGTAATTCATCAAGATGATGCCAGAGGAGCTGTGCATCGTCTGAGC 44A
Oy      92 AsnLyCySaSepleuThrThylsYsValVAlAsPyTrThrThrAlalysgluPha1a 111
Db      442 AACAAAGGGCATATGACGACGGAAGAAGATGATCCGTTCCGAAGACGAGAGACTTGACC 560
Oy      112 AspSerLeuGlyileProPhaleuGluThrSerAlAlaYsaMaLatrnAsValGlugla 133
Db      502 AGGAGATrAgGgtGTTCCTCTCGAGACCGACCGCCAAAGACTGCATGATGTGGAGTrrA 560
Oy      132 SerPheMetThrMeTaLaAlaGlulleylsYsArGmetCly 145
Db      562 GCCCTTCGTGGCTACGACCAAGAACTGAAATAACGGGAGCGG 603

RESULT 7
US-08-824-873-2
; Sequence 2, Application US/08824873
; Patent No. 5843717
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guglier, Karl
; TITLE OF INVENTION: NOVEL RAB PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,873
; FILING DATE: Filed Hewlett
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: pf-0240 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1340 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PANCNOT04
; CLONE: 738957
; US-08-824-873-2

Alignment Scores:
Prd. No.:          6,31e-29           Length:       1340
Score:             282.50              Matches:        70
Percent Similarity: 51.20%             ConservedAtts: 37
Best Local Similarity: 33.49%            Mismatches:     45
Query Match:       31.88%               Indels:         57
DB:                2                    gaps:          5

US-09-820-003B-2 (1-173) x US-08-824-873-2 (1-1340)
8 TrrAspyrleuPhelYsleuLeuIleGlyAspserGlyValGlyYsSerCysLeu 27

```

```

|||||
22 TACGACGTGCGCTTAAAGTCATGCTGGGGGAGCTGGGGGAGACCTGTCTG 81
QY Leu-LeuArgPheAlaAspAspThrTyr---ThrlGluSerTyrIleSerThrIleGlyVa 46
DB 82 CTGGGGTGGATTCAGAGATGCTGCTTCTCTGGGGGAGACCTTCATCTCCACCGTAGC-AT 140
QY 46 lAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGlnIle---- 64
DB 141 TGACTTCCGGAACAAAGTTCTGAGCTGATGTGTGAAGTGAAGCTGCACATGTGGGA 200
QY 64 ----- 64
DB 201 CACAGCTGTGACGAGCGGTTCCGAGTGTATCCATGCTACTACCGGGATGCTCAGTC 260
QY 65 -----GluSerPheAsnAsnValLysGlnTyrLe 74
DB 261 TCTGCTGCTGCTTACGATGTCACACAGAGCCCTTGTGACAACTCCAGGCTGTGCT 320
QY 74 uGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuValGlyAsnLysCy 94
DB 321 GACCCAGATCCACGAGTACGCCACAGACGCTGCGCTCATGCTGCGGAGAACAGGT 380
QY 94 sAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGluPheAlaAspSerIle 114
DB 381 GACACTCTGCCATGAGCGTGTGTGAAGAGAGAGAGAGCGGGGAGAGCTGGCCACAGAGTA 440
QY 114 uGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGlnSerPheMe 134
DB 441 TGGACTGGCCCTTCATGAGAGACAGCGCCACAGAGCGGCTTCACAGCTTGCCCTTAC 500
QY 134 tThrMetAlaAlaGluIleLysLysArgMetCylProGlyAlaThrAlaGlyAlaG1 154
DB 501 AGCCATACCAAGAGGTTGAGACAGCGC----- 528
QY 154 uLysSerAsnValLysIleGlnSerThrPro-----ValLysG1 167
DB 529 -----TCCATGAAGGCTCCACAGAGAGCGGCTTCCGGCTGCATGATTACGTTAAGAG 581
QY 167 nSerGlyGlyGly-----CysCys 173
DB 582 GGAGGGTCGAGGGGCGCTCTGCTGC 606

```

## RESULT 8

```

US-09-198-184-2
; Sequence 2, Application US/09198184
; Patent No. 6010859
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL RAB PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,184
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/824, 873
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

```

```

; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0240 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1340 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PANCNOT04
; CLONE: 738957
; US-09-198-184-2

```

## Alignment Scores:

Pred. No.:	6,31e-29	Length:	1340
Score:	282.50	Matches:	70
Percent Similarity:	51.20%	Conservative:	37
Best Local Similarity:	33.49%	Mismatches:	46
Query Match:	31.88%	Indels:	57
DB:	3	Gaps:	5

US-09-820-003B-2 (1-173) x US-09-198-184-2 (1-1340)

```

QY 8 TyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGlyValGlyLysSerCysLeu 27
DB 22 TACGACGTGCGCTTAAAGTCATGCTGGGGGAGCTGGGGGAGACCTGTCTG 81
QY 28 Leu-LeuArgPheAlaAspAspThrTyr---ThrlGluSerTyrIleSerThrIleGlyVa 46
DB 82 CTGGGGTGGATTCAGAGATGCTGCTTCTCTGGGGGAGACCTTCATCTCCACCGTAGC-AT 140
QY 46 lAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGlnIle---- 64
DB 141 TGACTTCCGGAACAAAGTTCTGAGCTGATGTGTGAAGTGAAGCTGCACATGTGGGA 200
QY 141 tThrMetAlaAlaGluIleLysLysArgMetCylProGlyAlaThrAlaGlyAlaG1 154
DB 501 AGCCATACCAAGAGGTTGAGACAGCGC----- 528
QY 154 uLysSerAsnValLysIleGlnSerThrPro-----ValLysG1 167
DB 529 -----TCCATGAAGGTCACAGAGAGCGGCTTCCGGCTGCATGATTACGTTAAGAG 581
QY 167 nSerGlyGlyGly-----CysCys 173
DB 582 GGAGGGTCGAGGGGCGCTCTGCTGC 606

```

## RESULT 9

```

US-08-773-423-4
; Sequence 4, Application US/08773423

```

```
US-09-820-003B-2 (1-173) x US-08-773-423-4 (1-847)
```

```
Patent No.5869291 ;  
GENERAL INFORMATION:  
APPLICANT : Hillman, Jennifer L.  
APPlicant : Goll, Surya K.  
Applicant : Bandman, Olga  
TITLE OF INVENTION: NOVEL RAB PROTEINS  
NUMBER OF SEQUENCES : 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304
```

```
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: IBM Compatible  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/773,423  
CLASSIFICATION: Herewith  
PRIORITY APPLICATION NUMBER:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0183 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TEXT:
```

```
INFORMATION FOR SEQ ID NO.: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 847 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMEDIATE SOURCE:  
LIBRARY: Consensus  
CLONE: Consensus  
US-08-773-423-4
```

```
Allment Scores:  
Pred. No.:  
Score: 1.81e-25 Length: 847  
Percent Similarity: 25.00 Matches: 66  
Best Local Similarity: 49.53% Conservative: 39  
Query Match: 31.13% Mismatched: 53  
DB: 28.78% Indels: 54  
Gaps: 5
```

```
OY 09-820-003B-2 (1-173) x US-08-773-423-4 (1-847)  
OY 7 GlutylAspTyrLeuPhelysLeuLeuIlelglYASPserGIyValIgLyssercys 26  
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||:  
Db 92 GATTAATAACTTTGGCTTCACAGTGCGTGCTGTGCAGATCCGAGTAGTGtgggAAGAACAA 151  
27 LeuluLeuArgPheAlaaspAsphTrlyrThrcLlSerTyrlLleserrlllegIyal 46  
||||| ||||| ::|||: |||||: |||||: |||||: |||||: |||||: |||||:  
Db 152 CTACTCTGCCGATTCACGCCCAATGAgTTcAcGACGACAGCGCACACCACCATCGGGGTT 211  
47 AspPhelylleAygrThrlllegluLeuaspLylysThrllLleserlnile----- 64  
:::||| |||||: |||: |||||: |||||: |||||: |||||: |||||:  
Db 212 GAGTTCACACCCGCACTGTGANtTTGGCCACCGCGTCGTCGTAAGCTCAAGATCTGGAC 271  
OY 64 ----- 64  
Db 272 ACAAGTCGGCTTGAGCGGTCACGACCACTACCTCGGCGTACTATCGTGGTCAGTGGGG 331  
OY 65 -----GlusErPheaSmaSnaVallYsgIntrrPLEU 74  
:::|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
```

```

Db      332  GCCCTCTGCTGTTTGAACCTAACCAAGCACAACACCTATGCTGTGTGGACCGATGCTG 391
Qy      75  GlnGluIAspAspArgYrAlaSerGluAsnValAsnLysLeuIleValGlyAsnLysCys  94
Db      332  :::::::::::::::::::::
Db      332  AAGAGAGCTATGACCATGCTGAAGCCACGATGCTGTCATCTGTGGTGAACAAAGT  451
Qy      95  AspleuThrThrLysLysValValAspYrThrThrAlaLysGluPheAlaAspSerLeu  114
Db      452  GACCTACGACGAGGGCCGAGAGTGCCACTGAGAGAGGCCGAATGTGGCTGTAACAACT  511
Qy      115  GlyLeuPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGlnSerPheMet  134
Db      512  GGAATCTCTCTCTGGAGACCTGACGCTCGACCTACCAATGCTGAGCTAGCTTGTAG  573
Qy      135  ThrMetAlaAlaGluIle-----LysLysArgMetGlyPro-----  146
Db      572  ACTGCTCTGAAAGAACTTTGGCAAGTGTCGCAAGACAGACAGAACACATCGGACC  633
Qy      147  --GlyAlaThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProVal  165
Db      632  AATCCATCATCTCTGGCGAGTGCCTGACGAGTGCAGAGTGCAGAGTGCAGAGTGCAG  666
Qy      166  LysGlnSerGlyGlyGly-----CysCys  173
Db      665  CAGGAGCGTGGCCCGGGGAGAAAGAGGCCCTGTGGC  700

RESULT 10
US-08-741-411-2
US-08-741-411-2
Sequence 2, Application US/08741411
Patent No. 6124116
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: NOVEL RAB PROTEINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/741,411
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0139 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-741-411-2

```

Alignment Scores:

Pred. No.:	2,966-22	Length:	848
Score:	231.50	Matches:	53
Percent Similarity:	45.45%	Conservative:	32
Best Local Similarity:	28.34%	Mismatches:	65
Query Match:	26.13%	Indels:	37
DB:	3	Gaps:	3

US-09-820-003b-2 (1-173) x US-08-741-411-2 (1-848)

```

OY 19 AspserglyValIglyLysserCysLeuLeuAlaaspPhaIaspPthryThrglu 38
DB 216 GACACTGGGGTGGGAATCAAGCATCGTGTGATTTGCCAGATCATTGACAC 275
OY 39 SerTrIleSerThrIleGlyValaspPheLysIleargThrIleGluLeuaspGlyLys 58
DB 276 AACATCAGCCCTACATTTGGGCATCTTTATGACCAAAACGTGCTTGGAATGAA 335
OY 59 ThrIleLysLeuGlnIle----- 64
DB 336 CTTCACAAGTTCCTCATCTGGGACACTGCTGTCAGAACGGTTTCATTCATTGGCTCC 395
OY 65 -----GluSer 66
DB 336 ATGTCATTCAGGCTCAGCTGACCTGTTATCGTGTATGATTAATTCACAGAGATTGA 455
OY 67 PheasnAsnValLysGlnTrpLeuGlnIleaspArgTrpAlaSerGluasnValasn 86
DB 456 TTTTACCTTGAAATAATGGGTCAAGAGCTGAAGAAACATGCTCCAGAAACATGTA 515
OY 87 LysLeuLeuValIglyLysLysCysaspLeuThrThrLysLysValValaspTrpThr 106
DB 516 ATGGCCATCGCTGGAACAGTGCACCTTCAGATATATGAGAGAGTTCCTGGAAGAT 575
OY 107 AlalysGluPheAlaSerSerLeuGlyIleProPheLeuGluThrSerAlaLysAla 126
DB 576 GCTAAGGATACGCTGAATCCATAGTGCATGGTGTGAACAAGTCAAAAATGCT 635
OY 127 ThrAsnValIgluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyPro 146
DB 636 ATTAATATCGAAGACCTCTTTCMA-----GGATAGCGCCGACAGATCCACCC 683
OY 147 GlyAlaThrAlaIglyLysIleGluLysSerAsnValLysIleGlnSerThrProValLys 166
DB 684 TTGACCCCGCCGAAATGGAACATGGAACATCAAGTGTGAGAG---CCAACCAATG 740
OY 167 GlnserGlyLysGlyCysCys 173
DB 741 CAAGCCAGCCCGGCTGCTGT 761

```

RESULT 11  
US-08-429-964-85  
Sequence 85, Application US/08429964  
Patent No. 5962243  
GENERAL INFORMATION:  
APPLICANT: BROWN, MICHAEL S.  
APPLICANT: GOLDSTEIN, JOSEPH L.  
APPLICANT: REISS, YUVAL  
APPLICANT: JAMES, GUY L.  
TITLE OF INVENTION: METHODS FOR THE IDENTIFICATION OF PARANEUTL  
TITLE OF INVENTION: TRANSFERASE INHIBITORS  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSER: ARNOLD, WHITE & DURKEE  
STREET: P.O. BOX 4433  
CITY: HOUSTON  
STATE: TEXAS  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/429,964  
FILING DATE: 27-APR-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/021,625  
FILING DATE: 16-FEB-1993  
CLASSIFICATION: 435  
APPLICATION NUMBER: US 07/822,011  
FILING DATE: ABANDONED  
CLASSIFICATION: 435  
APPLICATION NUMBER: PCT/US/91/02650  
FILING DATE: 18-APR-1991  
CLASSIFICATION: 435  
APPLICATION NUMBER: US 07/615,715  
FILING DATE: 20-NOV-1990  
CLASSIFICATION: 435  
APPLICATION NUMBER: US 07/510,706  
FILING DATE: 18-APR-1990 (ABANDONED)  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: PARKER, DAVID L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UTSD:432/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 85:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 607 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-429-964-85

Alignment Scores:

Pred. No.:	4,146-21	Length:	607
Score:	221.50	Matches:	55
Percent Similarity:	49.72%	Conservative:	35
Best Local Similarity:	30.39%	Mismatches:	56
Query Match:	25.00%	Indels:	35
DB:	2	Gaps:	4

US-09-820-003b-2 (1-173) x US-08-429-964-85 (1-607)

```

OY 12 PheLysLeuLeuLeuIleGlyaspSerGlyValIglyLysSerCysLeuLeuAlaPhe 31
DB 10 TATAAAGCTTGGTGGAGCTTGTGGCTGAGCAAGAGCTTCACGATACAGCTA 69
OY 32 AlaaspPthryThrGluSerTrpIleSerThrIleGlyValaspPheLysIlearg 51
DB 70 ATGACATCATTTTGTGACCAATATGATCCACAAATAGAGATCTCTACAG---AAG 126
OY 52 ThrIleGluLeuaspGlyLysThrIleLysLeuGlnIle----- 64
DB 127 CAAGTAGAATATGAGGAACACCTGCTCTTGATATTCGACACAGAGGTCAAG 186
OY 64 ----- 64
DB 187 GAGTACAGTCAATGAGGACGACGATACATGAGACTGGGAGCGCTTCTTGTGTATTT 246
OY 65 -----GluSerPheasnAsnValLysGlnTrpLeuGlnIleaspArg 79
DB 247 GCCATAAATATATCTAATCATTTGAAGATATTCACCTTATAGGAACATTAAGA 306
OY 80 TyrAla-----SerGluasnValasnLysLeuLeuValIglyLysLysCysaspLeuThr 98
DB 307 GTTAAGACTCTGAGAGGTACCTATGCTGCTAGTAGGAATAATATGATTTGCTCT 366
OY 99 LysLysValValaspTrpThrThrAlaLysGluPheAlaaspSerLeuGlyIleProPhe 118

```

[illegible]

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-29

```

Pred. No.:	3,34e-19	57/5
Score:	218.50	55
Percent Similarity:	49.11%	Matches:
Best Local Similarity:	30.39%	Conservative:
Query Match:	24.66%	Mismatches:
DB:	5	Indels:
		Gaps:
		4

US-09-820-003B-2 (1-173) x PCT-US93-06251-29 (1-5775)

OY		12	PheIysLeuLeuLeuLeuLeuLeuLeuAsperGlyAlaGlyLysSerCysLeuLeuLeuArgPhe	31
Dd		202	TATAAACTGTGGTACTTGAGACCTTCGTGGCGTAGCAAGAATGCCCTTACCATACACTA	261
OY		32	AlaAspAspThrTyrrThrgluSerrLytleSerthrilegilyalaasphelylearg	51
Dd		262	ATTCAACAATCATTTGTGGACGAATATGCATCCAAAATAGAGATTCCTCACAG--	AAG 318
OY		52	ThrIllegluleuaspolylysThrilleysleuglnlle-----	64
Dd		319	CANGATGTAATTAIGATGGAGAAACCCTGCTCTTGGATATTCTCGAACACAGAGTCAAAG	378
OY		64	-----	64
Dd		379	GAGTACAGTGCATGAGGAGCACCATACATGAGGAAGCTGGGAGGGCTTCTTGTGATTT	438
OY		65	-----glusErpheasnValylsgIntrrpleuglnIleasparg	79
Dd		439	GCCATAAATAATACATAATCATTTAAAGATATTCACCTATTAGAGAAACAATTMAAGA	498
OY		80	TyrAla---SercIusnvalalnslyrleuVallglyanlsyCyaspLeuThr	98
Dd		499	GTTAAGAGCATCGAAGATCTACCTATGTCCTAGTAGAATAATATGATTCCTCT	558
OY		99	LysIysValIyalnsPTrrThrAlalySGluPhelaasPerleuglyleProPhe	118
Dd		559	---AGAACAGTGTAGACACAAACAGGCTCAGGACTTACGCAACAGATTATGCAATTCCTTT	615
OY		119	IeuGlutThrSerAlalyssnlarthrasnalglungInserphemethMetAlala	138
Dd		616	ATTGANACATTCAGCAAGACAGACAGGCTGTGATGATGCCCTTCTATACATTAGTTCGA	675
OY		139	GIuIleLysArgmetcylProdlalathralagIylaglIysSerAsnVal	158
Dd		676	GAATTCGAAAAACATTAAGAAAAAGATGAGCAAAAGATGTTAAAAAGAAAAAGAACTCA	735
OY		159	Lys 159.	
Dd		736	AAG 738	
RESULT 14				
US-08-429-964-83				
Sequence 83 Application US/08429964				
Patent No. 5962243				
GENERAL INFORMATION:				
APPLICANT: BROWN, MICHAEL S.				
APPLICANT: GOLDSTEIN, JOSEPH L.				
APPLICANT: REISS, YUVAL				
APPLICANT: JAMES, GUY L.				
TITLE OF INVENTION: METHODS FOR THE IDENTIFICATION OF PARNESTYL				
NUMBER OF SEQUENCES: 85				
CORRESPONDENCE ADDRESS:				
ADDRESSEE: ARNOLD WHITE & DURREE				
STREET: P.O. BOX 4433				
CITY: HOUSTON				
STATE: TEXAS				

COUNTRY: UNITED STATES OF AMERICA  
ZIP: 77210  
COMPUTER READABLE FORM:

Alignment Scores:	
Pred. No.:	1,34e-20
Score:	217.50
Percent Similarity:	51.83%
Best Local Similarity:	31.71%
Query Match:	24.55%
DB:	2
Length:	574
Matches:	52
Conservative:	33
Mismatches:	44
Indels:	35
Gaps:	4

US-09-820-003B-2 (1-173) x US-08-429-964-83 (1-574)

OY		12	PheIysLeuIeuLeuIleuLglAspSerGlyValIGlYIlysScySylLeuIeuLeuPhe	31
OY		10	TATAAACTGTGGTACTTGAGACCTTGTGGCGTAGCAAGATGCCTTACCATCACACTA	69
OY		32	AlaAspAspThrTyTrThrGlusertyIleSerthrIleGlyValaSpPheIysIlearg	51
Db		70	ATTGACAATCAATTTTGTGGACGAATATGATCCAACAAATAGAGATTCCTACAG--AAG	126
OY		52	ThrIleGluLeuAspGelyIysThrIleIysLeuGlnIle-----	64
Db		127	CAGAATGTAATGATGAGAAACCCTGTCTCTTGATATTCGACACAGCAGTCAAAG	186
OY		64	-----	64
Db		187	GAGTACAGTGCATGAGGACCACTACATAGAGACTGGGAGGCGTTCTTTGTGTAATT	246
OY		65	-----GluSerPheAsnAsnValIysgIntPleuGlnGlnIleAspArg	79
Db		247	GCCATTAATTAATCTTAATCACTTTTAAGATATTCACATTTAGAGAAACAATTTAAAGA	306



```

0Y      80 TyrIala---SergIuaanValasnYlsLeuVaiGlyasNlysCysAspLeuThr 98
Db      307 GTTAAAGACTCTGAAGAATGTACCTTAGTGCTGTAGTGAATAAATGATTCGCCCTCT 366
0Y      99 LysIleValIaaPYrTrHrThraIalagLuphealaspSerLeuGIyleProPhe 118
Db      367 ---AGAACAGTGAACAACAAAACAGCGCTCAGAGCACTTGAACAAGATTGTAANTCCCTTT 423
0Y      119 LeuglThrsEralAlasnaIathrasValGlucInserPieMetRmetlaIala 138
Db      424 ATTCAAACATCATGCACAAACACAAACAGACAGAGTGGACAGATGCTTTTATCATTTGGTGACA 483
0Y      139 GluIleLysIys 142
Db      484 GAGATCGACACAA 495

RESULT 15
US-09-075-454-13
Sequence 13, Application US/09075454
Patent No. 6391580
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer I.
APPLICANT: Tang, Y. Tom
APPLICANT: Lal, Preeti
APPLICANT: Guglier, Karl J.
APPLICANT: Corley, Neil C.
APPLICANT: Patterson, Chandira
APPLICANT: Batra, Sajeev
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: RAS PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,454
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/766,551
FILING DATE: DECEMBER 12, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Geertone, Michael C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0168-1 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: OVARTU110
CLONE: Z703745
US-09-075-454-13

Alignment Scores:
Score: 2.24e-20 Length: 803
Pred. No.: 217.50 Matches: 60
Percent Similarity: 46.88% Conservative: 45

```

```

Best Local Similarity: 26.79% Mismatches: 61
Query Match: 24.55% Indels: 58
DB: 4 Gaps: 5
US-09-003B-2 (1-173) x US-09-075-454-13 (1-803)

OY 3 SerMetAspProGluTyrAspTyrLeu----- 11
Db 3 GAcTCTGCGCCCAACAGACCTGGTAATATACACGCTAAATCTTTTACCTACAGAG 62
OY 12 ---PhelysLeuLeuLeuIleGlyAspSerGlyValGlyLysSerCysLeuLeu-Ar 30
Db 63 GCTTACAAATATTGACTCTGGGACGCTGCATGCTGAGGAGATCTACTTCTCCATGAG 122
OY 30 pPheAlaAspAspThrTyrThrGluSerTyrLieserThrIleGlyAlaSpheAla 50
Db 123 ACTTTGCAAGAAATGAAATTCGAAAAAATATAGAGCCGACCCCTGGAGCTGATTCCAAAT 182
OY 50 eArGThrIleGluLeuAspGlyLysThrIleAlaLeuGluIle----- 64
Db 183 GAAACCCCTCATTTGGTGGTAGAGACGAACAACTTGTGACGTGGATACAGCTGTCA 242
OY 64 ----- 64
Db 243 GGAGAGATTCCAGAGTATGCCAAGTCTTACTTCCAGAAAGCAATGAGTGTTCCTCT 302
OY 65 -----GluSerPheAlaAspValGlyGlnTyrPheGluGluAla 78
Db 303 GTATGATGTTAATGTGAGAAAAGCTTCTTTACATTCAGATCAGATAGGTATATGATTGA 362
OY 78 pArGTYrAlaSerGlyAsnValAsnLysLeuLeuValGlyAsnLysCysAspLeuThr 98
Db 363 GGAATGACGCCCATTTAGACTGTCCCATTTATCTGTGTGAGAAACAAAGCCTACATTCGTGA 422
OY 98 r-----LysLysValAlaAspTYrThrAlaLysGluPheAla 112
Db 423 CACTGCTGCTACAGAGGACAAAAAGTGTCCAGGACACTTGGAGACAAAGTGGCCAT 482
OY 112 pSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln 132
Db 483 GACCTATGAGGCGATATTCTCTGTGAAACAAAGGCCAAAGATGGTCTTAACATATGTGAGGC 542
OY 132 rPheMetThrMetAlaIleAlaGluIleLysLysArgMetGlyProGlyAlaThrGly 152
Db 543 TGTCTTCACACTTGGCTGCGAAGATGAGAAAGAG-----ACTGACAGGA 587
OY 152 yAlaGluLysSerAsnValLysIleLysThrProValLysGlnSerCys-----Gl 170
Db 588 TGAAGAGATATCATTCACATCTAATACCGGGACCATTCACAAAAGTCCACACACATGAA 647
OY 170 yGlyCysCys 173
Db 648 GAATTGTTC 657

```

```

Length:      803
Matches:      60
Conservative: 45

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 22, 2003, 17:34:10 ; Search time 1419 Seconds  
(without alignments)  
1974.503 Million cell updates/sec

Title: US-09-820-003b-2  
Perfect score: 86  
Sequence: 1 MSSMNPEDYLFKLLIGDS.....EKSNVIGSTPVKSGGSCC 173

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+.p2n.model -DEV-xip  
-O/cgnt2.1/USPRO.spool/US09820003/unalc.18062003.144130.13104/app-query.fasta.1.327  
-DB-EST -OPMT-fastap -SUFFIX-p2n.rst -MINMATCH-0.1 -LOOPEXT-0 -LOOPEXT-0  
-UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST-45  
-DOCALLIGN-200 -NORM-ext -THR\_MAX-100 -THR\_MIN-0 -ALIGN-15 -MODE-LOCAL  
-OUTFMT-pio -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000  
-USPR-US09820003.CGNT.1.1.2874.gunat.18062003.144130.13104 -NCPV-6 -ICPV-3  
-NO\_MMAP -LARGEDECRY -NEG\_SCORES-0 -WAIT -DSPLOCK-100 -LONELIG  
-DEV\_TIMEOUT-120 -MAIN\_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6  
-FGAPEXT-7 -VGAPOP-10 -VGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :  
EST:  
1: em.estha.\*  
2: em.estha.\*  
3: em.estha.\*  
4: em.estha.\*  
5: em.estha.\*  
6: em.estha.\*  
7: em.estha.\*  
8: em.estha.\*  
9: em.estha.\*  
10: em.estha.\*  
11: em.estha.\*  
12: em.estha.\*  
13: em.estha.\*  
14: em.estha.\*  
15: em.estha.\*  
16: em.estha.\*  
17: em.estha.\*  
18: em.estha.\*  
19: em.estha.\*  
20: em.estha.\*  
21: em.estha.\*  
22: em.estha.\*  
23: em.estha.\*  
24: em.estha.\*  
25: em.estha.\*  
26: em.estha.\*  
27: em.estha.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860	97.1	948	AL539022	AL539022 AL539022
2	860	97.1	976	BM919860	BM919860 AGENCOURT
3	849	95.8	1130	BM542833	BM542833 AGENCOURT
4	847	95.6	862	BQ223273	BQ223273 AGENCOURT
5	847	95.6	944	BQ926234	BQ926234 AGENCOURT
6	831.5	93.8	646	BQ93078	BQ93078 AGENCOURT
7	825	93.1	957	BQ219698	BQ219698 AGENCOURT
8	822	92.8	864	BQ437146	BQ437146 AGENCOURT
9	820	92.6	901	BQ944368	BQ944368 AGENCOURT
10	819	92.4	1105	BM452262	BM452262 AGENCOURT
11	814	91.9	690	AM954378	AM954378 EST366448
12	809	91.3	945	AL530265	AL530265 AL530265
13	808.5	91.3	642	BQ073109	BQ073109 BQ073109
14	807	91.0	872	B1684554	B1684554 603306931
15	806	91.0	855	B1739625	B1739625 603361882
16	801	90.4	807	B6981507	B6981507 602660048
17	800	90.3	696	BM786746	BM786746 K-EST0065
18	800	90.3	890	BQ437514	BQ437514 AGENCOURT
19	795	89.7	673	B692590	B692590 602248943
20	795	89.7	846	B619962	B619962 602618161
21	794	89.6	1067	BM922791	BM922791 AGENCOURT
22	790	89.2	909	B6743352	B6743352 602634651
23	788.5	89.0	859	B1599480	B1599480 603246142
24	788.5	89.0	1034	BQ067968	BQ067968 AGENCOURT
25	788	88.9	611	A1048520	A1048520 hb61f06.x
26	788	88.8	828	B1652154	B1652154 603299164
27	783	88.4	838	B1858729	B1858729 603286319
28	783	88.4	913	BF970841	BF970841 602271483
29	780	88.0	933	BQ681067	BQ681067 AGENCOURT
30	778.5	87.9	588	BQ74071	BQ74071 BQ74071
31	777	87.7	901	BM450994	BM450994 AGENCOURT
32	777	87.7	937	B1738727	B1738727 603358761
33	776	87.6	676	AM913399	AM913399 uf52e09.y
34	775.5	87.5	612	AM959367	AM959367 fk37a09.y
35	775	87.5	1077	BM476507	BM476507 AGENCOURT
36	769.5	86.9	881	BQ424046	BQ424046 AGENCOURT
37	769	86.8	1047	BM800733	BM800733 AGENCOURT
38	768.5	86.7	643	AL535937	AL535937 AL535937
39	768.5	86.7	744	BM045935	BM045935 603624961
40	768.5	86.7	766	BM826700	BM826700 602750835
41	768.5	86.7	821	B1753331	B1753331 603026184
42	768.5	86.7	825	AL542911	AL542911 AL542911
43	768.5	86.7	826	BE561688	BE561688 60345726
44	768.5	86.7	837	AL520855	AL520855 AL520855
45	768.5	86.7	862	BQ941767	BQ941767 AGENCOURT

## ALIGNMENTS

RESULT 1  
AL539022  
LOCUS AL539022 LTI.F1013.Fbrn1 Homo sapiens cDNA clone COSDP303006 5  
DEFINITION Prime, mRNA sequence.  
ACCESSION AL539022  
VERSION AL539022.1 GI:12867866  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 948)  
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
Full-length cDNA libraries and normalization

## JOURNAL

Unpublished (2001)

COMMENT  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

source  
Location/Qualifiers

1. 948  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSODF030YG06"  
/dev\_stage="1st FL013.FBRn1"  
/dev\_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"  
/lab\_host="DH10B"  
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 277 a 187 c 239 g 244 t 1 others  
ORIGIN

## Alignment Scores:

Pred. No.: 1.87e-94 Length: 948  
Score: 860.00 Matches: 173  
Percent Similarity: 84.39% Conservative: 0  
Best Local Similarity: 84.39% Mismatches: 0  
Query Match: 97.07% Indels: 32  
DB: 9 Gaps: 1

US-09-820-003b-2 (1-173) x AL539022 (1-948)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLeuLeuLeuIleGlyAspSer 20  
DB 166 AAGTCACAGCAATGAAATCCGAAATATGATATTTATTCAGTACTCTGATGCGACATCA 225  
QY 21 GlyValIleGlySerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
DB 226 GGGGTGGAAAGCTCTGCTCTCTTGTAGCTTGACATGATACATATACAAAGCTTAC 285  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 286 ATCAGCACAATTTGGTGGATTTCAAAATAGAACTATAGATTAGACGGGAAACAAATTC 345  
QY 61 LysLeuGlnIle----- 64  
DB 346 AAGCTTCAAAATATGGGACACAGCAGCAAGAAAGATTTGCAACATCACCCTCAGTTAT 405  
QY 65 -----GluSerPheAsn 68  
DB 406 TACAGAGAGCCCATGCATCATCTGTTGATGATGACATGACAGAGTCCCTTCAT 465  
QY 69 AsnValLysGlnTyrPheGlnGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
DB 466 AATGTTAAACAGTGGCTCGAGAAATAGATGCTTATGCGACAGGAAATGTCACAAATTC 525  
QY 89 LeuValIleGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThalLys 108  
DB 526 TTGCTAGAGGAACAAATGATCTGACCAACAAAGAAAGTATGATACACACACAGCAAG 585  
QY 109 GluPheAlaAspSerLeuGlyTyrLeuPheLeuGlnTyrSerAlaLysAsnAlaThrAsn 128  
DB 586 GAATTTGCTGATCCCTGGATATCCGTTTGGAAACAGGCTGTAACAAATGCAACGAT 645  
QY 129 ValGluGlnSerPheMetThrMetAlaIleGlnIleLysLysArgMetGlyProGlyAla 148  
DB 646 GTAGAACAGTCTTTATATGACATGCGACCTGAGATTAAAGCGAATGGGTCGCCGAGACA 705

QY 149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
DB 706 ACAGCTGGTGGTGGTGAAGATCCCAATGTAAATTCAGACACATCCAGTCAAGACATCA 765  
QY 169 GlyGlyGlyCysCys 173  
DB 766 GGTGGAGGTGCTGTC 780

## RESULT 2

BM919860  
LOCUS

DEFINITION BM919860 976 bp mRNA linear EST 12-MAR-2002  
AGENCOURT.6708189 NIH\_MGC\_120 Homo sapiens cDNA clone IMAGE:5749635  
5', mRNA sequence.

ACCESSION BM919860  
VERSION BM919860.1 GI:19370239  
KEYWORDS EST.

SOURCE human.  
ORGANISM

REFERENCE Homo sapiens  
AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE NIH-MGC http://mhc.nci.nih.gov/.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1992)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLNL12779 row: f column: 04  
High quality sequence stop: 757.

## FEATURES

source

1. 976  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5749635"  
/clone\_11b="NIH\_MGC\_120"  
/lab\_host="DH10B"  
/note="Organ: pooled pancreas and spleen. Vector: pCMV-SPORT; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb. Insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research genetics tracking code 025. Note: this is a NIH\_MGC Library."

BASE COUNT 285 a 196 c 246 g 248 t 1 others  
ORIGIN

## Alignment Scores:

Pred. No.: 1.95e-94 Length: 976  
Score: 860.00 Matches: 173  
Percent Similarity: 84.39% Conservative: 0  
Best Local Similarity: 84.39% Mismatches: 0  
Query Match: 97.07% Indels: 32  
DB: 14 Gaps: 1

US-09-820-003b-2 (1-173) x BM919860 (1-976)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLeuLeuLeuIleGlyAspSer 20  
DB 177 AAGTCACAGCAATGAAATCCGAAATATGATATTTATTCAGTACTCTGATGCGACATCA 236  
QY 21 GlyValIleGlySerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
DB 237 GGGGTGGAAAGCTCTGCTCTCTTGTAGCTTGACATGATACATATACAAAGCTTAC 296  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60

```

Db      297 ATCAGACACATTGGTGGATTTCAAAATAGACTATAGAGTTAGACGGAAACATC 356
Oy      61 LysleuGlnIle-----64
Db      357 AACGTTCAAAATATAGGACACAGACAGCCAGAAAGTTTCACATCACCCTCTGATTAT 416
Oy      65 -----GluserPheasn 68
Db      417 TACAGAGAGCCCGCAGCATCATGTGTCGTATGATGTCAGACATCAGAGATCAGATTCAT 476
Oy      69 AsnValIysGlnITrPleuGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88
Db      477 AATGTTAAACAGTGGCTCAGAAATGATCGTATGCGCAAGTAAATGTCACAAATG 536
Oy      89 LeuValIglYAsnLysCysAspLeuThrThrLysValValAspTyrThrAlaLys 108
Db      537 TTGGTAGGAGAAATGATGATCTGACACAAAGAAAGTAGACACACACAGCGAAG 596
Oy      109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnLathrasn 128
Db      597 GAATTTGCTGATTCCTCGAATTCGTTTGGAAACAGCTGTAAGAAATGCAACGAT 656
Oy      129 ValGluInserPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148
Db      657 GTAAACAGCTTTCATGACATGACAGCTGAGATTAAAAAGCATGGGTCCGAGACA 716
Oy      149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168
Db      717 ACAGCTGGTGGCTGAGAACTCCAAATGTTAAATTCAGACACCTCCAGTCAAGCATCA 776
Oy      169 GlyIleGlyCysCys 173
Db      777 GTGAGAGTTGCTGC 791

RESULT 3
BM542833 1130 bp mRNA linear EST 20-FEB-2002
LOCUS BM542833
DEFINITION AGENCOURT_6426035 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5520670
5', mRNA sequence.
ACCESSION BM542833
VERSION BM542833.1 GI:18772697
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mhc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strusberg, Ph.D.
Email: rstrubs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILML at:
http://image.llnl.gov
Plate: ILAM12185 row: a column: 23
High quality sequence stop: 597.
Location/Qualifiers
1. 1130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5520670"
/clone_1lb="NIH_MGC_71"
/tissue_type="leukosarcoma"
/lab_host="DH108 (phage-resistant)"
/notes="Organ: uterus; Vector: PCMV-SPOK6; Site: 1; Ncti:
Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2.1 kb."

```

```

BASE COUNT 319 a 241 c 290 g 280 t
ORIGIN
Alignment Scores:
Pred. No.: 5.4e-93 Length: 1130
Score: 849.00 Matches: 171
Percent Similarity: 83.41% Conservative: 0
Best Local Similarity: 83.41% Mismatches: 2
Query Match: 95.82% Indels: 32
DB: 13 Gaps: 1
US-09-820-003b-2 (1-173) x BM542833 (1-1130)
Oy      1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20
Db      178 ATGTCACAGCATGAATCCCGAAATATGATTATTATTCAAAGTACTTCATATGGCGACTCA 237
Oy      21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40
Db      238 GGGGTTGGAAAGTCTTCCTCTCTTAGGTTTCACATATACATATACAGAAAGCTAC 297
Oy      41 IleSerThrIleGlyValaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60
Db      298 ATCAGACAAATGGGTGGATTTCAAAATAGAACTTAGAGTTAGACGGAAACAAATC 357
Oy      61 LysleuGlnIle-----64
Db      358 AACCTTCAAAATATGGGACACAGACAGCCAGAAAGATTTCGAACATCACCTCCAGTTAT 417
Oy      65 -----GluserPheasn 68
Db      418 TACAGAGAGCCCATGCGCATCATGTTGTATGATGTCAGACATCAGAGATCCTTCATAT 477
Oy      69 AsnValIysGlnITrPleuGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88
Db      478 AATGTTAAACAGTGGCTCAGAAATAGATCGTATGCGCAAGTAAATGTCACAAATG 537
Oy      89 LeuValIglYAsnLysCysAspLeuThrThrLysValValAspTyrThrAlaLys 108
Db      538 TTGGTAGGAGAAATGATGATGACACCAAAAGAAAGTAGTACTACACACAGCGAAG 597
Oy      109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnLathrasn 128
Db      598 GAATTTGCTGATTCCTCGAATTCGTTTGGAAACAGCTGTAAGAAATGCAACGAT 657
Oy      129 ValGluInserPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148
Db      658 GTGAAACAGCTTTCATGACATGACAGCTGAGATTAAAAAGCATGGGTCCGAGACA 717
Oy      149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168
Db      718 ACAGCTGGTGGCTGAGAACTCCAAATGTTAAATTCAGACACCTCCAGTCAAGCATTC 777
Oy      169 GlyIleGlyCysCys 173
Db      778 AGTGAAGTTGCTGC 792

RESULT 4
BQ223273 862 bp mRNA linear EST 02-MAY-2002
LOCUS BQ223273
DEFINITION AGENCOURT_7518264 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:606561
5', mRNA sequence.
ACCESSION BQ223273
VERSION BQ223273.1 GI:20404673
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mhc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```



Db		200	AATGAGACAGCAGCGCCAGGAAGAATTTCGAAACAATCACTTCAGATTATACAGAGGA	259
Oy		65	-----GusierPheasnantValys	71
Db		260	GCCCATGCATCATCAGTGTGTGTATCATGTGCACAGATCAGAGACTCTCTTCATTAACCTTTAA	319
Oy		72	GIATPTPLengInguilileasparfyrThlaSercJuaenVaaanYleauYleuValgY	91
Db		320	CATGTGCTTCACAGAGNATGATGCTTCAGCCAGAGNAATGATCAACAGATTGTGGTAGGG	379
Oy		92	AsnIysCyJaapLeuthrThrIlyslYsVaYaIaAsPyTrYthrThralalysGIuPhueHa	111
Db		380	AAACAATGTGACCTTGACCACCAAAGAAAGTGTGACTCACACACAGGAAGAAATTGCCA	439
Oy		112	ASPSerLengIylleProPheLengIunThSerAlalySaenalathraNaenValdugln	131
Db		440	GATTTCCCTTGGAAATTCATTTTTTGGAAAACAGTGTAAAGAACCGACAGAAATATACACAG	499
Oy		132	SerPheMetthrMetaIaalaGIuIlleLyalyatgMetglYproGlyalathralagly	151
Db		500	TCTTTATGATAGAGATGAGCAGCATATTAATAACGAAATGGGTCTCTGACACTTCACACTGGT	559
Oy		152	GIYAIAIGluLYSSerAsnaVallysIIeqlSerThrProvalLYSGInserIyGlY	111
Db		560	GGTGGCGAGAAAGTCCATGTATTAATCCAGACACTCCAGTCAAGCAGTCAAGTGGAGGC	619
Oy		172	CYSaCS	173
Db		620	TCCTGC	625
RESULT 6				
Bj093078				
LOCUS				
DEFINITION			646 bp mRNA linear EST 12-DEC-2000	
ACCESSION			Bj093078 NIBB Mochii normalized Xenopus early gastrula library	
VERSION			Bj093078	
KEYWORDS			Bj093078.1 GI:17592487	
SOURCE			EST.	
ORGANISM			African clawed frog.	
TITLE			Xenopus laevis	
JOURNAL			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
COMMENT			Amphibia; Batrachia; Anura; Mesobatrachia; Pipelidae; Pipidae;	
REFERENCE			Xenopodinae; Xenopus.	
AUTHORS			1 (bases 1 to 646)	
			Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara	
			Y.	
			Expressed genes in X. laevis embryo	
			Unpublished (2001).	
			Contact: Tadasu Shin-I	
			Center For Genetic Resource Information	
			National Institute of Genetics	
			1111 Yata, Mishima, Shizuoka 411-8540, Japan	
			Tel: 81-559-81-6856	
			Fax: 81-559-81-6855	
			Email: tshini@genes.nig.ac.jp.	
FEATURES			Location/Qualifiers	
source			1..646	
			/organism="Xenopus laevis"	
			/db_xref="taxon:8355"	
			/clone="Xl141fl8"	
			/clone_lib="NIBB Mochii normalized Xenopus early gastrula	
			library"	
			/tissue_type="whole embryo"	
			/dev_stage="Stage 10.5"	
			/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; CDNAS	
			were oligo-dt primed and directionally cloned. Staging	
			according to Nieuwkoop and Faber. Library is subcloned	
			and was constructed by N. Garrett and A.M. Zorn,	
			(Wellcome/CRC Institute). "	
BASE COUNT			208 a 129 c 161 g 148 t	
ORIGIN				

[illegible]





```

OY 69 AsnValylsGlnTrpLeuGlnIleAspArgTyrAlaSerGluAsnValAsnLysIleu 88
DB 484 AATGTTAAACAGTGGTCAGAGAAATAGATCGTTATGCCAGTGAATATGCACAAATTTG 543
OY 89 LeuValGlyAsnLysGlyAspLeuThrThylsValValAspTyrThrThAlaLys 108
DB 544 TTGGTAGGAGCAACATGATGTGACACAAAGATGAGCTACACACACAGAGAG 603
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128
DB 604 GAATTCGTGATTCCTTGGAATTCCTTTTGAAACCAAGCTAAGAAATGCACAGAT 663
OY 129 ValGluInserPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148
DB 664 GTAGACAGCTCTTTCATTCAGTCAGTCAGATTAATAAAGCAATAGCGTCCGAGACA 723
OY 149 ThrAlaGlyLysAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168
DB 724 ACAGCTGGTGTCTGAGAAATGCATTAATTCAGACACTCCAGTCAAGCAGTCA 783
OY 169 Gly-GlyGlyCys 172
DB 784 GGTGGAGAGTTGC 796

RESULT 9
BO944368 901 bp mRNA linear EST 21-AUG-2002
LOCUS BO944368
DEFINITION IMAGE:6441874 5', mRNA sequence.
ACCESSION BO944368
VERSION BO944368.1 GI:22359846
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 901)
NIR-MGC http://mgc.nci.nih.gov/.
Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT
AUTHORS Email: cgabbs-remail.nih.gov
TITLE Tissue Procurement: Gilbert Smith, Ph.D.
JOURNAL CDNA Library Preparation: Life Technologies, Inc.
Unpublished (1999) DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
COMMENT DNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AMJ3965 row: m column: 11
High quality sequence stop: 662.
FEATURES
source location/Qualifiers
1 901
/organism="Mus musculus"
/strain="FVB/N-3"
/DB_xref="taxon:10090"
/clone="IMAGE:6441874"
/clone_11b="NCL_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORT6; Site:1: Saliv;
Site:2: Notti; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by: Life Technologies. Investigator
providing samples: Gilbert Smith, NIR"
BASE COUNT 246 a 197 c 243 g 214 t 1 others
ORIGIN
Alignment Scores: 1,34e-89 Length: 901
Pred. No.: 820.00 Matches: 170
Score: 82.618 Conservative: 1
Percent Similarity:

```

```

Best Local Similarity: 82.13% Mismatches: 2
Query Match: 92.55% Indels: 34
DB: 14 Gaps: 1
US-09-820-003b-2 (1-173) x BO944368 (1-901)
OY 1 MetSerMetLaspProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20
DB 188 ATGTCAGCATGAATCCGAATATGATTTATTCAGATTCTTGATTTGGCGATTCT 247
OY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40
DB 248 GGGGTGGAAATCTCTGCTTCTCTTACGTTTGCACATATACGTAATCGGAAAGCTAC 307
OY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60
DB 308 ATCAGCACAAATTTGTGATTCATACGATACGACTATACGTTAGATCGGAAAGCATC 367
OY 61 LysLeuGlnIle----- 64
DB 368 AAGCTACAGATATGGGACACAGCAGGCCAGAAAGATTTCGAACATCTCCAGTTAT 427
OY 65 -----GluSerPheAsn 68
DB 428 TACAGAGAGCCCATGCGCATATGTTGTTATGATGTGACACATCAGAGTCCCTCAAT 487
OY 69 AsnValylsGlnTrpLeuGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88
DB 488 AAGCTTAAACAGTGGTCAGAGAAATAGATCGTTATGCCAGTGAATATGCACAAATTTG 547
OY 89 LeuValGlyAsnLysGlyAspLeuThrThrLysLysValValAspTyrThrThAlaLys 108
DB 548 TTGGTAGGAGCAACAAATGTGACCTGACCAAGAAAGATGATACACACACAGCAAG 607
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128
DB 608 GAATTCGAGATTCCTTGGAATTCATTTTGGAAACCAAGTCTTAAAGACCAACAAAT 667
OY 129 ValGluInserPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148
DB 668 GTAGACAGCTCTTTCATTCAGTCAGTCAGATTAATAAAGCAATAGCGTCCGAGAGCT 727
OY 149 ThrAlaGlyLysAlaGluLysSerAsnValLysIleGlnSerThrProVal-LysAlaLys 168
DB 728 ACAGCTGGTGTGGCGGAGAAATGCCATTTAAATTCAGAGAACCTCCAGTCAAGCAGC 787
OY 168 ergLysGlyGlyCysCys 173
DB 788 CAGGTGAGAGCTGCTGC 804

RESULT 10
BM452262 1105 bp mRNA linear EST 05-FEB-2002
LOCUS BM452262
DEFINITION AGENCOURT_6386191 NIR-MGC_72 Homo sapiens cDNA clone IMAGE:5526535
5', mRNA sequence.
ACCESSION BM452262
VERSION BM452262.1 GI:18501302
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1105)
NIR-MGC http://mgc.nci.nih.gov/.
Unpublished (1999) National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT
AUTHORS Email: cgabbs-remail.nih.gov
TITLE Tissue Procurement: ATCC/DCID/DTF
JOURNAL CDNA Library Preparation: Life Technologies, Inc.
Unpublished (1999) DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
COMMENT DNA distribution: MGC clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LINT at:  
 http://image.llnl.gov  
 plate: L1AM12200 row: j column: 12  
 High quality sequence stop: 733.  
 Location/Qualifiers  
 1..1105

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NIB\_MGC\_72"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Skin; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 2 kb. Library constructed by Life  
 Technologies."

BASE COUNT 320 a 216 c 279 g 282 t 8 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 2,43e-89 Length: 1105  
 Score: 819.00 Matches: 172  
 Percent Similarity: 83.50% Conservative: 0  
 Best Local Similarity: 83.50% Mismatches: 1  
 Query Match: 92.44% Indels: 34  
 DB: 13 Gaps: 1

US-09-820-003b-2 (1-173) x BM452262 (1-1105)

1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 182 ATGTCACGATGATATCCGAAATATGATTTATTCACATTCCTGATGGCGACTCA 241  
 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 242 GGGGTGGAAAGCTTGCTCTCTTCTAGCTTGCAGATGATACATATACAGAACTAC 301  
 41 IleSerThrIleGlyValAlaPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
 302 ATCAGCACAAATGGTGTGATTTCAAAATAGACATATAGATAGACGCGAAACAAATC 361  
 61 LysLeuGlnIle----- 64  
 362 AAGCTTCAATATATGGACACAGCAGCGCAAGAAATTTGCAACATCACTCCAGTTAT 421  
 65 -----GluSerPheAsn 68  
 422 TACAGAGAGAGCCCATGGCATCATAGTGTGATGATGTGACAGATCAGAGAGTCCCTCAAT 481  
 69 AsnValLysGlnTyrPheGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
 482 AATGTTAAACACTGCTGCGAGGAAATAGATCCTTATGCCAGTCAAAATGTCAAAATATG 541  
 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
 542 TTGCTAGGAGAACAAATGTGATCTGACACCAAGAAAGTAGTACACCAACAGGGAAG 601  
 109 GluPheAlaAspSerLeuGlyTyrProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
 602 GAATTTGGCGATTCCTTGGATTCCTGTTTGGAAACAGTGCATAGAAATGCACAGAT 661  
 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
 662 GTAGAACAGTCTTTCATGACATGACAGCTGAGATTTAAACCAATGGGTCGCCGAGACA 721  
 149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 722 ACAGCTGGGTGGTGTGAGAA-TCCAATGTTAAATTCAGAGACATCCAGTCAAGACATCA 780  
 169 -GlyGlyLysCysCys 173  
 781 NGGTGAGGTTCCTGC 796

RESULT 11  
 AM954378 690 bp mRNA linear EST 01-JUN-2000  
 LOCUS EST166448 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.  
 DEFINITION AM954378  
 ACCESSION AM954378.1 GI:8144061  
 VERSION EST.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 1 (bases 1 to 690)  
 Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt  
 I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and  
 Quackenbush,J.  
 Assessment of gene expression patterns in a model of colon tumor  
 metastasis using a 19,200 element cDNA microarray

TITLE  
 JOURNAL  
 COMMENT  
 Contact: John Quackenbush  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 3528  
 Fax: 301 838 0208  
 Email: john@etlgr.org  
 Plate: 63  
 Seq primer: Reverse.

FEATURES  
 source  
 1..690  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="MAGE resequences, MAGE"  
 /note="Vector: pBluescriptSKm"

BASE COUNT 219 a 130 c 168 g 173 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.81e-89 Length: 690  
 Score: 814.00 Matches: 169  
 Percent Similarity: 83.25% Conservative: 0  
 Best Local Similarity: 83.25% Mismatches: 1  
 Query Match: 91.87% Indels: 33  
 DB: 10 Gaps: 1

US-09-820-003b-2 (1-173) x AM954378 (1-690)

1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 31 ATGTCACGATGATATCCGAAATATGATTTATTCACATTCCTGATGGCGACTCA 90  
 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 91 GGGGTGGAAAGCTTGCTCTCTTCTAGCTTGCAGATGATACATATACAGAACTAC 150  
 41 IleSerThrIleGlyValAlaPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
 151 ATCAGCACAAATGGTGTGATTTCAAAATAGACATATAGATAGACGCGGAAACAAATC 210  
 61 LysLeuGlnIle----- 64  
 211 AAGCTTCAATATATGGACACAGCAGCGCAAGAAATTTGCAACATCACTCCAGTTAT 270  
 65 -----GluSerPheAsn 68  
 271 TACAGAGAGAGCCCATGGCATCATAGTGTGATGATGTGACAGATCAGAGAGTCCCTCAAT 330  
 69 AsnValLysGlnTyrPheGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
 331 AATGTTAAACACTGCTGCGAGGAAATAGATCCTTATGCCAGTCAAAATGTCAAAATATG 390  
 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
 391 TTGCTAGGAGAACAAATGTGATCTGACACCAAGAAAGTAGTACATACCAACAGCAGAG 450

QY 109 GluphalaaspserleuglylleprophelaugluThrsAlaYsaAnaIarPha 128  
 DB 451 GAAATTCCTATTCCTTCGGAATTCCTTTTGCAACCAAGTCAATGCAACGAAAT 510  
 QY 129 ValgluInserPhemethrMetAlaAlaGlulIleYlsArGmetGlyProGlyAla 148  
 DB 511 GTAAGACAGCTTCATGACATGACGACGAGATTAAGCAATGAGTCCGCGACCA 570  
 QY 149 ThrAlaGlyAlaGluYlsSerAsnValYlsIleGlnSerThr-ProValYlsGlnSe 168  
 DB 571 ACACGGGGTGTGCTCCGGAATTCATGTAATTAATTCAGACATCTTCAGTCAACGAC 630  
 QY 168 IGlyGly 170  
 DB 631 AGGTGGG 637  
 RESULT 12  
 LOCUS AL530265 945 bp mRNA linear EST 13-FEB-2001  
 DEFINITION AL530265 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CSDD009YH19 5  
 prime, mRNA sequence.  
 ACCESSION AL530265  
 VERSION AL530265.1 GI:12793758  
 KEYWORDS EST  
 SOURCE human  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 945)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.  
 FEATURES  
 source  
 1. 945  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CSDD009YH19"  
 /clone\_1lb="LTI\_NFL001\_NBC4"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer, five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed  
 by life technologies. Contact : Peng Huang Life  
 Technologies, a division of Invitrogen 9800 Medical Center  
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
 8371 Email : liliang@life.com  
 http://fulllength.invitrogen.com"  
 BASE COUNT 280 a 186 c 224 g 246 t 9 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3.19e-88 Length: 945  
 Score: 809.00 Matches: 171  
 Percent Similarity: 83.01% Conservative: 0  
 Best Local Similarity: 83.01% Mismatches: 2  
 Query Match: 91.31% Indels: 34  
 DB: 9 Gaps: 1  
 US-09-820-003b-2 (1-173) x AL530265 (1-945)  
 QY 1 MetSerSerMetAsnProGluTyrAspTyrIleuphelysleuleuileuilegIYAspser 20  
 DB 121 ATGTCCAGCATGAAATTCGGAATATGATTAATTAATTCAGTTCGATTCGCGACCA 180  
 QY 21 GlyValGlySerCysleuLeuLeuAphnaIaAspThrTyrThrGluSerTyr 40

DB 181 GGGGTGGAAAGTCTGTGCTTCCTTCAGTTCAGATATACATACAGAACTCTAC 240  
 QY 41 TlaserThrIleGlyValaAspPheYlsIleArgThrIleGluLeuAspGlyYsrThrIle 60  
 DB 241 ATCAGCAATATGGTGTGGATTCCAAATAGAACTATAGAGTACGCGGAAACAAATC 300  
 QY 61 LysLeuGlnIle----- 64  
 DB 301 AAGCTTCAATATGCGACACGACGCGCAGGAAAGATTCGAACATCACTCCACTAT 360  
 QY 65 -----GluSerPheAsn 68  
 DB 361 TACAGAGAGCCCATGCATCATAGTGTGTATGATGTGACAGATCAAGATCCTTCAT 420  
 QY 69 AsnValYlsGlnTTPleuGlnGluIleAspArgTyrIleAspSerGluAsnValaYlsIleu 88  
 DB 421 AATGTTAAACAGTGTGCGAGGAATAGATGTTATCCAGTGAATAATGTCACAAATTC 480  
 QY 89 LeuValGlyAsnYlsCysAspIleuphThrIleYlsValYlsValaAspTyrThrThrAlaYls 108  
 DB 481 TTGCTAGGGAACAAATGTGATCTGACCAACAAAGAAATAGTACTACACACAGCAAG 540  
 QY 109 GluphalaaspserleuglylleprophelaugluThrsAlaYsaAnaIarPha 128  
 DB 541 GAATTCGCTGATTCCTTCGGAATTCGTTTGGAAACAGTCTAGAAAGCAACGCAAT 600  
 QY 129 ValgluInserPhemethrMetAlaAlaGlulIleYlsArGmetGlyProGlyAla 148  
 DB 601 GTGAAGACATCTTCATGACGATGACGACGTAGATTAAGCAATGGGTCCGGA-GCA 659  
 QY 149 ThrAlaGlyAlaGluYlsSerAsnValYlsIleGlnSerThr-ProValYlsGlnSer 168  
 DB 660 ACAGCTGGGGTGTGCGAGGAAGTCAATGTTAAATTCAGACATCTCCAGTCAACGACATCA 719  
 QY 169 -GlyGlyGlyCysCys 173  
 DB 720 GGGTGGAGGTGTCTGC 735  
 RESULT 13  
 LOCUS B073109 642 bp mRNA linear EST 11-DEC-2001  
 DEFINITION B073109 N18B Mochli normalized Xenopus tailbud library Xenopus  
 laevis cDNA clone X1108p01 5', mRNA sequence.  
 ACCESSION B073109  
 VERSION B073109.1 GI:17503298  
 KEYWORDS EST  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 Xenopodinae; Xenopus.  
 REFERENCE 1 (bases 1 to 642)  
 AUTHORS Kiteyama, A., Teresaka, C., Mochli, M., Deno, N., Shin-I, T. and Kohara  
 Y.  
 TITLE Expressed genes in X. laevis embryo  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Tadasu Shin-1  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 FEATURES  
 source  
 1. 642  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="X1108p01"  
 /clone\_1lb="N18B Mochli normalized Xenopus tailbud  
 library"  
 /tissue\_type="whole embryo"  
 /dev\_stage="stage 25"

/note="Vector: pBSRN3; Site.1: NotI; Site.2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieuwkoop and Faber. Library is subtracted and was constructed by N. Garrett and A.M. Zorn, (Wellcome/CRC Institute)."

BASE COUNT 212 a 137 c 150 g 142 t 1 others

## ALIGNMENT SCORES:

Alignment Scores: 2.03e-88 Length: 642  
Pred. No.: 808.50 Matches: 164  
Percent Similarity: 81.68% Conservative: 4  
Best Local Similarity: 81.19% Mismatches: 33  
Query Match: 91.25% Indels: 2  
Gaps: 13

US-09-820-003b-2 (1-173) x BJ073109 (1-642)

QY 4 MetlanpPogluTYrAspTYrLeuPhelLysLeuLeuileGlyAspSerGlyValGly 23  
DB 10 ATGATATCCCGGANTATGACTATTTATTCATAATTTGATTTGATGAGACTGTGTGCGA 69  
QY 24 LysSerCysLeuLeuLeuArgPheAlaAspAspThrTYrThrGluSerTYrIleSerThr 43  
DB 70 AGCTTGTCTTCTGCTTGGCTTGCATGACATGACACAGAAAGTTACTACATCAGTACA 129  
QY 44 IleGlyValAspPheLysIleArgTYrIleGluLeuAspGlyLysThrIleLysLeuGln 63  
DB 130 ATGCTGTGTATTTAAATCAGACGACTATGAGTTAGATGAGAAACAACTGCA 189  
QY 64 Ile----- 64  
DB 190 ATTTGGGAGACTGCCGCCAGAGAAAGATTTCGAACCATTTACATCAAGTTACTACAGGGA 249  
QY 65 -----GluSerPheAsnValLys 71  
DB 250 GCCCAGCGCATATTGTGTGTATGACGTACACAGACGAGAACTCTTACACATTCAG 309  
QY 72 GlnTYrLeuGlnGluIleAspArgTYrAlaSerGluAsnValAsnLysLeuValGly 91  
DB 310 CAGTGGCTTCAGAAATGATCGTTATGCCAGTCAAGCTTAACAGTTATTCGTAGGC 369  
QY 92 AsnLysCysAspLeuThrThrLysValValAspTYrThrAlaLysGluPheAla 111  
DB 370 AACCAATGTATCTCACACAAAGAAAGTTGTGACATCACACAGCAAGAGATTTC 429  
QY 112 AspSerLeuGlyIleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnValGluGln 131  
DB 430 GACCTCTGGGAATTCATTTTGGAAACAAAGCGAAAGCGACAAAGCTAGACAG 489  
QY 132 SerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThrAlaGly 151  
DB 490 GCGCTTCAAGCAAAAGACAGCGAGATCGAAAGCGAAAGCGCCCTGGCCACACAGC 549  
QY 152 GlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGlyGly 171  
DB 550 GGTCAAGGAGAG---AACGTCAAAATCCAGAGCACTCCAGTCAGTACGTAGTGAGGA 606  
QY 172 CysCys 173  
DB 607 TGCTGC 612

RESULT 14  
BI684554 872 bp mRNA linear EST 18-SEP-2001  
LOCUS 603306931F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5343151 5',  
DEFINITION mRNA sequence.  
ACCESSION BI684554  
VERSION BI684554.1 GI:15647170  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 872)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC Clone distribution information can be found through the I.M.A.G.E. Consortium/LMN at:  
<http://image.llnl.gov>  
Plate: ILAM11871 row: m column: 08  
High quality sequence stop: 802.

FEATURES  
source location/qualifiers  
1..872  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5343151"  
/clone\_1lb="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 235 a 184 c 246 g 207 t

## ORIGIN

US-09-820-003b-2 (1-173) x BI684554 (1-872)

Alignment Scores: 4.94e-88 Length: 872  
Pred. No.: 807.00 Matches: 172  
Score: 83.09% Conservative: 0  
Percent Similarity: 83.09% Mismatches: 35  
Best Local Similarity: 91.08% Indels: 1  
Query Match: 13 Gaps: 1

QY 1 MetSerSerMetAsnProGluTYrAspTYrLeuPhelLysLeuLeuileGlyAspSer 20  
DB 206 ATGTCACACATGAAATCCGGAATATGATTTATTTCAAGTTACTGATTCGGGATTC 265  
QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTYrThrGluSerTYr 40  
DB 266 GGGGTGGAAAGTCCCTGCTTCTCTTAGGTTGCAGATGATACGTATACGAAAGCTAC 325  
QY 41 IleSerThrIleGlyValAspPheLysIleArgTYrIleGluLeuAspGlyLysThrIle 60  
DB 326 ATTCAGCAAAATTTGGTGTGATTTCAAGTACGAATATAGATTAAGTGGAAACATTC 385  
QY 61 LysLeuGlnIle----- 64  
DB 386 AAGCTACAGATATGGGACACACAGAGCCGAGAAAGATTTCGAACAATCACTTCAGTTAT 445  
QY 65 -----GluSerPheAs 68  
DB 446 TACAGAGAGCCCATGTCATCAGTTGTGTATGATGACAGATCAGAGTCCATCA 505  
QY 68 nasnValLysGlnTYrLeuGlnGluIleAspArgTYrAlaSerGluAsnValAsnLysLe 88  
DB 506 TAACTTAAACAGTGGCTGCGAGAGATGATCGCTAGCCAGTGAATATGCAACACTT 565  
QY 88 uLeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTYrThrThrAla 108  
DB 566 GTTGCTAGGAGAAACAATGTGACCTGACACCAAGAAAGAACTAGTACGCTACACAGCAA 625  
QY 108 sGluPheAlaAspSerLeuGlyIleProPheLeuGlnThrSerAlaLysAsnAlaThrAs 128

Db 626 GGAATTGACAGATTCCTTGGATTCATTTTGGAAACAGCTGTCAAGCAACGAA 685  
OY 128 nvaIgluIserPheMetThrMetAlaIaIuIleIysIysArgMet-GlyProGlyA 148  
Db 686 GTGAAACACTCTTTCATGACGATGACGATGATTAAGGAAATGGGCTCCTGAG 745  
OY 148 IatThrIaIaIyIaIuIyIaIuIyS-SerAsnValIyIleIginSerThrProValIySglu 167  
Db 746 CTACAGCTGTGTGTCGCCGAGAAAGCTCCAAAGTTAAATCCAGACCTCCATGACAG 805  
OY 168 SerGlyIyIyIyCys 172  
Db 806 TTAGGTGAGGCTGC 820  
RESULT 15  
LOCUS B1739625 855 bp mRNA linear EST 20-SEP-2001  
DEFINITION 603361882F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:536876 5',  
mRNA sequence.  
ACCESSION B1739625 GI:15716638  
VERSION B1739625  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 855)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: rgs@gsf.fritz.fritz.nih.gov  
Tissue Procurement: The Cepko Laboratory  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at:  
http://image.llnl.gov  
Plate: L14M11938 row: d column: 21  
High quality sequence stop: 849.  
Location/Qualifiers  
1. 855  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:536876"  
/clone\_lib="NIH\_MGC\_94"  
/issue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1; NotI;  
Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."  
BASE COUNT 234 a 181 c 233 g 205 t 2 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 6,36e-88 length: 855  
Score: 806.00 Matches: 169  
Percent Similarity: 82.218 Conservative: 2  
Best Local Similarity: 81.258 Mismatches: 35  
Query Match: 90.978 Indels: 13  
Gaps: 1  
US-09-820-003b-2 (1-173) x B1739625 (1-855)  
OY 1 MetSerSerMetLeuAnProGluTyrAspTyrLeuPheIySleuLeuIleGlyaspSer 20  
Db 185 ATGTCCAGCAGATGAAATCCCAATATGATTATTATTCAGTTACTTCGATGCGATTCT 244  
OY 21 GIYValGIyIySerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40

Db 245 GGGATTGAAATCCTGCTTCTCTCTAGCTTGCAGATGATACGATATGCAAGCTAC 304  
OY 41 IISerThrIleGlyValaIaIaPheIySleuIleArgThrIleGluLeuAspGlyIySThrIle 60  
Db 305 ATCGACAAATTTGGTGTGATTTCAAGATACGAACTATAGTAGATGGAAACAAATC 364  
OY 61 IyLeuGluIle----- 64  
Db 365 AAGCTACAAATATGAGACACAGACAGCCAGAAAGATTTTGAAACAATCTACTTCAGATT 424  
OY 65 -----GluSerPheAsn 68  
Db 425 TACAGAGAGCCCATGCAATCATAGTTGTATGATGTGACAGATCAGAGTCTCTCATAT 484  
OY 69 AsnValIySgluIleIlePheIySleuIleAspArgTyrAlaSerGluAsnValaIySleu 88  
Db 485 AAGCTTAAACAGTGTGCTGACAGAGATGATGATGATGATGATGATGATGATGATGAT 544  
OY 89 LeuValaIyAsnIySAspLeuThrIySValaIyAspTyrThrAlaIyS 108  
Db 545 GTGGTAGCGAACAATGTGACCTGACCAAAAGTAAGTACACACACAGCAAG 604  
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluIlePheSerAlaIySAsnAlaIyS 128  
Db 605 GAAATTGACAGATTCCTTGGATTCATTTTGGAAACAGCTGTAGACGACGAAAT 664  
OY 129 ValaIgluIserPheMetThrMetAlaIaIuIleIySAspArgMetGly--ProGlyA 148  
Db 665 GTGAAACACTCTTTCATGACGATGACGATGATTAAGGAAATGGGCTCCTGAG 724  
OY 148 IatThrIaIaIyIaIuIyS-SerAsnValIyIleIginSerThrProValIySglu 168  
Db 725 CTACAGCTGTGTGTCGCCGAGAAAGCTCCAAAGTTAAATCCAGACCTCCATGACAG 784  
OY 168 erGly-GlyGlyIyCysCys 173  
Db 785 CAGGTGAGGCTGCTGC 802

Search completed: June 22, 2003, 18:55:41  
Job time: 1427 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 22, 2003, 16:26:39 ; Search time 216 Seconds  
(without alignments)

1803.684 Million cell updates/sec

Title: US-09-820-003b-2

Perfect score: 866  
Sequence: 1 MSSNMPDYLYFKLLIGDS.....EKSNNKIQSTPVKGGGCC 173

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Command line parameters:  
MODEL=frame\_p2n\_model -DEV=klp  
-O=/cgn2/\_1/USPTO-spool/US09820003/runat\_18062003\_144129\_13082/app\_query.fasta\_1.3227  
-DB=N-Geneseq\_101002 -QPM=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOFCUT=0  
-LOOEXT=0 -ONITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdl  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09820003.ecgn.1.1.396.eRunat\_18062003\_144129\_13082 -NCPU=6 -ICPU=3  
-NO\_MMAP -LANG=ENGLISH -NCG\_SCORES=0 -WAIT -DSPBLLOC=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860	97.1	723	ABK83978	Human cDNA differe
2	860	97.1	2528	ABK60009	Human secreted pro
3	768.5	86.7	925	AAZ27332	Human Rab protein,
4	768.5	86.7	939	AAZ21661	Human breast and o
5	768.5	86.7	1898	AA944924	cDNA encoding nove
6	768.5	86.7	1944	AA193456	Human polynucleoti
7	733.5	82.8	1965	AA534668	Human DNA for a no
8	683	77.1	1791	ABL05983	Drosophila melanog
9	657.5	74.2	1202	AAFL14053	Aspergillus oryzae
10	639	72.1	959	AAAC33987	Arabidopsis thaila
11	633	71.4	932	AAAC35200	Arabidopsis thaila
12	624	70.4	1193	AAAC32684	DNA encoding novel
13	619.5	69.9	777	AAAC42684	Arabidopsis thaila
14	619.5	69.9	881	ABN98761	Arabidopsis thaila
15	612	69.1	666	AAAC42764	Arabidopsis thaila
16	612	69.1	932	AAAC48274	Arabidopsis thaila
17	597	67.4	607	AAH87926	Arabidopsis thaila
18	580	65.5	594	AAAC49581	peppermint plant o
19	570	64.3	1380	ABF07578	Arabidopsis thaila
20	529	59.7	5019	ABF05982	Arabidopsis thaila
21	460	51.9	749	ABFL14316	Fusarium venenatum
22	452	51.0	615	AAH87928	Drosophila melanog
23	444	50.1	911	AAAC38429	Trichoderma reesei
24	442	49.9	1101	AAAC34080	peppermint plant o
25	440.5	49.7	1203	AAAC34685	Arabidopsis thaila
26	431	48.6	1129	AAAC44482	Arabidopsis thaila
27	429.5	48.5	705	AAAC49208	Arabidopsis thaila
28	428.5	48.4	1023	AAAC51491	Arabidopsis thaila
29	428.5	48.4	1025	AAAC34347	Arabidopsis thaila
30	419	47.3	1866	ABAO9160	Human rab8 homolog
31	417	47.1	2411	ABO54410	Human ovarian anti
32	414	46.7	1161	AAH75182	Nucleotide sequenc
33	414	46.7	2497	AAAC60878	Human cancer agent
34	414	46.7	2497	AAAC60884	Human cancer agent
35	414	46.7	2497	AAAC60893	Human cancer agent
36	414	46.7	2497	AAAC60895	Human cancer agent
37	414	46.7	2497	ABV25781	Human prostate exp
38	414	46.7	2497	ABV30037	Human prostate exp
39	414	46.7	3077	AAH13912	Human cDNA sequenc
40	413	46.6	866	AAH04301	Human cDNA clone (
41	408	46.0	2247	ABF29661	Drosophila melanog
42	405	45.7	1492	AAAC81336	Zea mays DNA fragm
43	405	45.7	559	AAAC44411	Arabidopsis thaila
44	403.5	45.5	716	AAAC40104	Human Rab10 cDNA
45	403.5	45.5	861	AAAC40108	Human Rab10 cDNA

## ALIGNMENTS

RESULT 1  
ABK83978  
ID ABK83978 standard; cDNA: 723 BP.

XX ABK83978;

DT 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #549.

XX Human: ss; granulocytic cell; DNA chip; bacterial infection;  
XX viral infection; parasitic infection; protozoal infection;  
XX fungal infection; sterile inflammatory disease; postrasis;  
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;  
XX adult respiratory distress syndrome; inflammatory bowel disease;  
XX Crohn's disease; ulcerative colitis; peridontal disease;  
XX granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.  
OS  
XX  
XX  
PM MO200228999-A2.  
PN  
XX  
XX  
PD 11-APR-2002.  
PE  
XX 03-OCT-2001; 2001WO-US30821.  
PF  
XX 03-OCT-2000; 2000US-237189P.  
PR  
XX  
XX  
PA (GENE-) GENE LOGIC INC.  
PI  
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
PI  
XX WPI; 2002-435328/46.  
DR  
XX  
XX  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity  
PS  
XX  
XX  
XX Claim 1; SEQ ID No 549; 114pp; English.

The invention relates to detecting (M1) granulocyte (GC) activation by (GCA), by detecting the level of expression of gene(s) (Gs) identified by DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

**SQ** Sequence 723 BP; 225 A; 142 C; 180 G; 176 T; 0 other;

Alignment Scores:	
Pred. No.:	4,336-89
Score:	860.00
Percent Similarity:	84.39%
Best Local Similarity:	84.39%
Query Match:	97.07%
DB:	24
Length:	723
Matches:	173
Conservative:	0
Mismatches:	0
Indels:	32
Gaps:	1

US-09-820-003B-2 (1-173) x ABK83978 (1-723)

OY			MesSerSerMetIAsnProGluTyrAspTyrLeuPheLysLeuLeuLeuIleGlyAspSer	20
Db			51 ATGTCCAGCGATGAATCCCGAATATGATTATTATTCAAGTAACCTTGATGGCGACTCA	111
OY			21 GtValGIyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrTrpGluSerTyr	40
Db			111 GGGGTGGAAAGTCCTTCCTTTCTTAGAGTTTGCGAGATATATCATATACAGAAGCTAC	17
OY			41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle	60
Db			171 ATCGGCACAATGGGTGGATTTCCAATAATGAACATATAGAGTTAGACGGGAAAAACAATC	233
OY			61 LysLeuGlnIle-----	64
Db			231 AAGCTTCAAAATATGGGACACAGCAGGCCAGGAAAAGATTTCCGAACAAATCACCTCCAGTAT	29
OY			65 -----GluSerPheasn	68
Db			291 TACAGAGAGCCCATGGCATCATAGTTGTGTATGATGATGCACAGATCAGAGAGTCTTCAT	35
OY			69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu	88
Db			351 AATGTTAAACAGTGGCGTCGACGAAATAGATCGTTATGCCAGTGAATAATGCACAAATATG	411
OY			89 LeuValGIyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys	10
Db			411 TTGGTAGGGAACAAATGTATGTACGCCACCACAAAGAAAGTATGATACATCACACACGGCAAG	47
OY			109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn	12
Db			471 GAATTTCTGATTCCTCGTGAATTCGGTTTTTGGAAACACAGTCTGAAGATGCACGAAC	53
OY			129 ValGIuGlnSerPheMetThrMetAlaIleGluIleLysLysArgMetGlyProGlyAla	14
Db			531 GTGGAACAGCTTTTCATGACGATGGCAGCGCTGACATTTAAAAAGCGAATGGGTCCCGGAGCA	59
OY			149 ThrAlaGlyValAlaGluLysSerAsnValLysIleGluSerThrProValLysGlnSer	16
Db			591 ACAGCTGGTGGTGTGGAAGTCCAAATGTTAAATTCAGAGCACCTCAGTCAACAGACTCA	65
OY			169 GlyGIyGIyCysCys	173
Db			651 GTGGAGAGTTGCTGC	665
RESULT 2				
AAC60009				
ID	AAC60009	standard; cDNA; 2528 BP.		
XX				
AC	AAC60009;			
XX				
DT	26-JAN-2001	(first entry)		
XX				
DE		Human secreted protein gene 44 SEQ ID NO:54.		
XX				
KW		Human; secreted protein; diagnosis; cytostatic; immunosuppressive;		
KW		nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;		
KW		antidiabetic; antiinflammatory; antifiber; vulnereary; anticonvulsant;		
KW		antibacterial; antifungal; antiparasitic; cardiant; gene therapy;		
KW		cancer; immune disorder; cardiovascular disorder; wound healing;		
KW		neurological disease; infectious disease; chromosome identification; ss.		
XX				
OS	Homo sapiens.			
XX				
WO200058356-A1.				
PN				
PD	05-OCT-2000.			
XX				
PF	22-MAR-2000; 2000MO-US07535.			
XX				
PR	26-MAR-1999; 99US-0126511.			
PR	17-DEC-1999; 99US-0172413.			
XX				
VA	(HUMA-) HUMAN GENOME SCI INC.			

(HUMA-) HUMAN GENOME SCI INC.







Oy		22	VaIGILySseScYslLeuleleuarphellaAspRfThrTYrrThgIusertYrile	41
Db		124	GtGGCGCAATCATGCTCCTCCGCCGTTTCTGTAGTAGACAGTCACAGAGACTCATC	189
Oy		42	SeTThrIIeGIyValAAsPheLySlLaRghTrIIeGIuLeAAsPgLYkStPhrIIeLys	61
Db		184	AGCACCATCGSGRGTRGCATTCAAGATCCGAACATCGACACTGSATGGCCAAACTATCAA	243
Oy		62	LeuGInIle-----	64
Db		244	CITTCAGATCTGGACACACGGGGCCAGAGACGGTCCGACCATCTCCAGCTACTAC	303
Oy		65	-----giUserpheasn	69
Db		304	CGGGGGCGTCATGCATCATCTGTGTATGACGTCACTACACAGGAATCTACGCCAAC	363
Oy		70	ValIyGGLnTPLeuGInGuIlIeaSPArTyAlASerGIuaNValAaNYSLeuleu	89
Db		364	GTFACACATGTGGCTGCAGAGATGATACCGCTATGCCAGGAAACGTCAATAAGTCTTG	423
Oy		90	ValGILyAaNLyScYsApLeuRfThrThrySLySaIVaIAsPyTYrrThrIIaIySGlu	109
Db		424	GTFGGCAACAAGACGACCTCAACCAAGAAGSTGTGTGCAACACACAGCAAGAG	483
Oy		110	PheAlaAspSerIeuGIlyIleProPheuGlThrSerAlIySaNaIARhnaVal	128
Db		484	TFTGCAGACTCTGGGCATCCCCCTTCTTGSAAGCAAGCCCAAMATGCAACACTGC	543
Oy		130	GIuGlnserPheMetThrMetAlaLaclUIeLySLySaRgmEtIyProGIyAlaThr	149
Db		544	GAGACAGGCTTCATGACCAATGAGCTGCTCAATCAAAAACGGAATGGGCTCGAGACCC	603
Oy		150	AlaGILyAlaIaGTLySerrAnValIySLIeGInSerThrProValIySGInserGIly	169
Db		604	TCGGGGGGC---GAGCGGCCCAATCTCAAGATGACACGACCCCTGTAAAGCCGCTGGC	660
Oy		170	GIlyIyScYcYS	173
Db		661	GGTGGCTGTGGC	672
RESULT 5				
ID	AA544924		standard; cDNA; 1898 bp.	
XX	AA544924;			
DF	18-DEC-2001	(first entry)		
DE	cDNA encoding novel human secretory protein, Seg ID No. 5.			
XX	Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;			
KM	Ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;			
KM	transgenic animal; Alzheimer's disease; Parkinson's disease; burn;			
KM	amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;			
KM	ulcer; osteoporosis; bone degenerative disorder; periodontal disease;			
KM	gut protection; lung; liver fibrosis; immune deficiency; infection;			
KM	severe combined immunodeficiency; SCID; autoimmune disorder; allergy;			
KM	multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;			
KM	fertility; analgesic; pain; antigen; ss.			
OS	Homo sapiens.			
PN	MO200166689-A2.			
PD	13-SEP-2001.			
PF	05-MAR-2001; 2001WO-US04942.			
PR	07-MAR-2000; 2000US-0519705.			
PR	19-MAY-2000; 2000US-0574454.			
PR	17-JUN-2000; 2000US-0536193.			
R	14-JUL-2000; 2000US-0616847.			

[illegible]

```

Db 166 AGCACATCGGGGTGACTTCAAGATCCGACCATCGAGCTGATGCGAAATATATCAAA 225
Qy 62 LeuGlnIle----- 64
Db 226 CTTGAGATCTGGGACACAGCGGGCCAGAGAGGTTCCGACCATCTCCAGCTACTAC 285
Qy 65 -----GluserPheasnAsn 69
Db 286 CGGGGGGCTCAGGCGATCATCGTGTATGACGTGACTGACACGAAATCTTACCCCAAC 345
Qy 70 VallysgIntPLeuGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89
Db 346 GTGAAGCACTGGCTCAGAGATTACCGCTATGCCAGCGAGAACTCAATAGCTCCCTG 405
Qy 90 ValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrAlaLysGlu 109
Db 406 GTGGGCAACAGAGAGCGCTCCACCAAGAAAGTGTGGACACACACACAGCGAGAG 465
Qy 110 PheAlaAspSerLeuGlyIleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnVal 129
Db 466 TTTCGAGACTCTCTGGGATCCCTCTTGGAGACGCGCCAAAGATGCCCAATGTC 525
Qy 130 GluGlnSerPheMetThrMetAlaAlaGlnIleLysLysArgMetGlyProGlyAlaThr 149
Db 526 GAGCAGGCGCTCATGACCATGCTGCTCTCAATCAAAAGCGGATGGGCTGGAGCAGCC 585
Qy 150 AlaGlyAlaGlnLysSerAsnValLysLysIleGlnSerThrProValLysGlnSerGly 169
Db 586 TCTGGGGGC---GAGCGGGCCCAATCTCAAGATCGACACACGCCCTGTAAAGCGGCTGGC 642
Qy 170 GlyGlyCysCys 173
Db 643 GTGGCTGTGTC 654

RESULT 6
AAI93456
ID AAI93456 standard; cDNA, 1944 BP.
AC AAI93456;
XX
XX 06-NOV-2001 (first entry)
DE Human polynucleotide SEQ ID NO 13516.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AAO13525.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX
XX Claim 1; SEQ ID NO 13516; 1399pp + Sequence Listing; English.
XX

```

```

CC The invention relates to human polynucleotides (AAI9941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC treatment/inhibitin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
SO Sequence 1944 BP; 372 A; 618 C; 552 G; 402 T; 0 other.

Alignment Scores:
Pred. No.: 4.72e-78 Length: 1944
Score: 768.50 Matches: 156
Percent Similarity: 80.39% Conservative: 8
Best Local Similarity: 76.47% Mismatches: 7
Query Match: 86.74% Indels: 33
DB: 22 Gaps: 2

US-09-820-003B-2 (1-173) x AAI93456 (1-1944)
Qy 2 SerSerMetLanProGlnTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGly 21
Db 56 GCGGCGATGAACCCCGAATATGACTACCTGTTTAAAGCTGTTGATGGGCGGATCAAGC 115
Qy 22 ValGlyLysSerCysLeuLeuLeuArgPheAlaAspArgTyrThrGlnSerTyrIle 41
Db 116 GTGGGCAAGTATGCTGCTCTCTGCTGCTTCTATGACAGTACAGAGAGCTACTATC 175
Qy 42 SerThrIleGlyValAspPheLysIleArgThrIleGluAsnArgLysThrIleLys 61
Db 176 AGCACATCGGGGTGAGCTTCAGATCCGAAACATCGAGTGGATGCAAAATATATCAA 235
Qy 62 LeuGlnIle----- 64
Db 236 CTTGAGATCTGGGACACAGCGGCCGAGACGGTTCCGACCATCTCCAGTACTAC 295
Qy 65 -----GluserPheasnAsn 69
Db 296 CGGGGGGCTCAGTGCATCATGCTGTGTATGACGTACACTGACAGGAATCTTACGCCAAC 355
Qy 70 VallysgIntPLeuGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89
Db 356 GTGAAGCAGTGGCTGACAGAGATTGACCGCTATGCCAGCGAGAACTCAATTAAGCTCTTG 415
Qy 90 ValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrAlaLysGlu 109
Db 416 GTGGGCAACAGAGAGCGACTCACCACCAAGAGTGTGGCAACACACACAGCGAGAG 475
Qy 110 PheAlaAspSerLeuGlyIleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnVal 129
Db 476 TTTCGAGACTCTCTGGGATCCCTCTTGGAGAGAGCGCCAAAGAAATGCAACAAATGTC 535
Qy 130 GluGlnSerPheMetThrMetAlaAlaGlnIleLysLysArgMetGlyProGlyAlaThr 149
Db 536 GAGCAGGCGCTCATGACCATGCTGCTCTCAATCAAAAGCGGATGGGCGCTGGAGCAGCC 595
Qy 150 AlaGlyAlaGlnLysSerAsnValLysLysIleGlnSerThrProValLysGlnSerGly 169
Db 596 TCTGGGGGC---GAGCGGGCCCAATCTCAAGATCGACACACGCCCTGTAAAGCGGCTGGC 652
Qy 170 GlyGlyCysCys 173
Db 653 GTGGCTGTGTC 664

RESULT 7
AAS34668

```

ID	AA34668 standard; DNA; 1965 bp.	PR	08-SEP-2000; 2000US-0231414.
XX		PR	08-SEP-2000; 2000US-0232060.
AC	AA34668;	PR	08-SEP-2000; 2000US-0232061.
XX		PR	12-SEP-2000; 2000US-0231968.
DT	17-DEC-2001 (first entry)	PR	14-SEP-2000; 2000US-0232397.
XX		PR	14-SEP-2000; 2000US-0232398.
DE	Human DNA for a novel foetal antigen, SEQ ID No 2092.	PR	14-SEP-2000; 2000US-0232399.
XX		PR	14-SEP-2000; 2000US-0232400.
KM	Human: foetal tissue antigen; ds; antinflammatory; neuroprotective;	PR	14-SEP-2000; 2000US-0232401.
KM	immunomodulator; cardiovascular; cytostatic; nephrothropic;	PR	14-SEP-2000; 2000US-0233063.
KM	cardiovascular; autoimmune disease; rheumatoid arthritis;	PR	14-SEP-2000; 2000US-0233064.
KM	hyperproliferative disorder; breast neoplasm; cancer;	PR	14-SEP-2000; 2000US-0233065.
KM	cardiovascular disorder; cardiac arrest; cerebrovascular disorder;	PR	21-SEP-2000; 2000US-0234223.
KM	cerebral ischaemia; angiogenesis; nervous system disorder;	PR	21-SEP-2000; 2000US-0234224.
KM	Alzheimer's disease; infection; ocular disorder; corneal infection;	PR	25-SEP-2000; 2000US-0234997.
KM	wound healing; epithelial cell proliferation; food additive.	PR	25-SEP-2000; 2000US-0234998.
XX		PR	26-SEP-2000; 2000US-0235484.
OS	Homo sapiens.	PR	27-SEP-2000; 2000US-0235834.
XX		PR	27-SEP-2000; 2000US-0235836.
PM	WO200155312-A2.	PR	29-SEP-2000; 2000US-0236327.
XX		PR	29-SEP-2000; 2000US-0236367.
XX		PR	29-SEP-2000; 2000US-0236368.
PD	02-AUG-2001.	PR	29-SEP-2000; 2000US-0236369.
XX		PR	29-SEP-2000; 2000US-0236370.
PF	17-JAN-2001; 2001WO-US01321.	PR	02-OCT-2000; 2000US-0236802.
XX		PR	02-OCT-2000; 2000US-0237037.
XX		PR	02-OCT-2000; 2000US-0237038.
PR	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000; 2000US-0237039.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000; 2000US-0237040.
PR	24-FEB-2000; 2000US-0184664.	PR	13-OCT-2000; 2000US-0239935.
PR	02-MAR-2000; 2000US-0186350.	PR	13-OCT-2000; 2000US-0239937.
PR	16-MAR-2000; 2000US-0189874.	PR	20-OCT-2000; 2000US-0240960.
PR	18-MAR-2000; 2000US-0190076.	PR	20-OCT-2000; 2000US-0241785.
PR	19-MAR-2000; 2000US-0198123.	PR	20-OCT-2000; 2000US-0241786.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000; 2000US-0241787.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000; 2000US-0241808.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000; 2000US-0241809.
PR	07-JUL-2000; 2000US-0216880.	PR	01-NOV-2000; 2000US-0241826.
PR	11-JUL-2000; 2000US-0217487.	PR	08-NOV-2000; 2000US-0246474.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000; 2000US-0246475.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000; 2000US-0246538.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000; 2000US-0246609.
PR	14-AUG-2000; 2000US-0225759.	PR	08-NOV-2000; 2000US-0246610.
PR	18-AUG-2000; 2000US-0226279.	PR	08-NOV-2000; 2000US-0246611.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000; 2000US-0249207.
PR	22-AUG-2000; 2000US-0226686.	PR	17-NOV-2000; 2000US-0249208.
PR	23-AUG-2000; 2000US-0227182.	PR	17-NOV-2000; 2000US-0249209.
PR	30-AUG-2000; 2000US-0227009.	PR	17-NOV-2000; 2000US-0249210.
PR	01-SEP-2000; 2000US-0228924.	PR	17-NOV-2000; 2000US-0249211.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000; 2000US-0249212.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000; 2000US-0249213.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000; 2000US-0249215.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000; 2000US-0249216.
PR	05-SEP-2000; 2000US-0229409.	PR	17-NOV-2000; 2000US-0249217.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000; 2000US-0249218.
PR	05-SEP-2000; 2000US-0230437.	PR	17-NOV-2000; 2000US-0249219.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000; 2000US-0249244.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000; 2000US-0249245.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000; 2000US-0249264.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000; 2000US-0249265.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000; 2000US-0249297.



CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116175-AB130511) expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB157217-AB172072)  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [http://wipo.int/pub/publ/publseq\\_sequences](http://wipo.int/pub/publ/publseq_sequences).

XX Sequence 1791 BP; 579 A; 379 C; 360 G; 473 T; 0 other;

Alignment Scores:  
Pred. No.: 2,6e-68 Length: 1791  
Score: 683.00 Matches: 143  
Percent Similarity: 75.24% Conservative: 12  
Best Local Similarity: 69.42% Mismatches: 17  
Query Match: 77.09% Indels: 34  
DB: 23 Gaps: 3

US-09-820-003b-2 (1-173) x AB105983 (1-1791)

OY 1 MetSerSerMetAspProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
DB 252 ATGTCATCTGTGTAATCGGAATACGATATCTTAAAGTGGCTTATGAGACTCG 311  
OY 21 GlyValGlyLysSerCysAlaLeuLeuAlaAspAspThrTyrGluSerTyr 40  
DB 312 GCGGTGGAAAGTCGTCCTCTCTGTCGCGGACATACATACAAAGATAT 371  
OY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 372 ATCGACCAATCGAGTGGATTTCAAAATCAGACTATAGACATCGAAGAACATT 431  
OY 61 LysLeuGlnIle-----GluSerPheAsn 64  
DB 432 AAATCGCAATTCGGAGTACTGTCGGCAGACCGCTTCCGACATCAGCTTCTATAT 491  
OY 65 -----GluSerPheAsn 68  
DB 492 TATAGGGGCGCCCATGCGATCATCTTCGTATACGATTCACGACGACGAGCTTTCAC 551  
OY 69 AsnValLysGlnIlePheGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
DB 552 AATGTGAAGCAGTGGCTGAGAGATCGAGCGGTATGCTGTAGAAATGTCAACAAGTTG 611  
OY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValAlaSerTyrThrAlaLys 108  
DB 612 CTGGTTGGCAACAGGCGACTTACCCACAGAGAGTAGTCGACACACACAGCTGCG 671  
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
DB 672 GAGTACCGCCCGCCAGTGGCATTCCTCTTGAACCTTGGCCAAAGGCCCCACAC 731  
OY 129 ValGlnIleSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
DB 732 GTTAGAGAGCGCTTCATGATGATGCGCGGAGATCAAGAAATCCCTCGCGCGCGTCC 791  
OY 149 ThrAlaGlyLysAlaGluLysSerAsnValLysIle-----GlnSerThrProValLysGln 167  
DB 792 AGCCGCACTGAC---AACGTCAGCAAGTAGAAATGATCAAGAGAGCTCAGTGAAGAAC 848  
OY 168 SerGlyLysGlyCysCys 173  
DB 849 ACCAAATCCGCTGCTGC 866

RESULT 9  
AAFI4053  
ID AAFI4053 standard; cDNA; 1202 BP.  
XX  
AC AAFI4053;

XX 13-MAR-2001 (first entry)  
XX Aspergillus oryzae EST SEQ ID NO:6576.  
DE  
XX Multiple gene expression; filamentous fungal cell; EST;  
XX Aspergillus oryzae; Fusarium venenatum; Aspergillus niger;  
XX Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
XX culture condition; environmental stress; spore morphogenesis;  
XX metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
XX Aspergillus oryzae.  
XX  
XX WO200056762-A2.  
XX  
XX 28-SEP-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07781.  
XX  
XX 22-MAR-1999; 99US-0273623.  
XX  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
XX (NOVO ) NOVO NORDISK AS.  
XX  
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
XX WPI: 2000-594572/56.  
XX  
XX Monitoring differential expression of genes in filamentous fungal cells  
XX uses fluorescence-labeled nucleic acids isolated from the cells and a  
XX substrate of expressed sequence tags -  
XX  
XX Claim 88; Page 2691; 3161pp; English.  
XX  
XX The present invention describes a method for monitoring differential  
XX expression of genes in a first filamentous fungal (FF) cell relative to  
XX expression of the same genes in one or more second filamentous fungal  
XX cells. The method uses fluorescence-labeled nucleic acids isolated from  
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
XX are used in the methods for monitoring differential expression of genes  
XX in a first filamentous fungal (FF) cell relative to expression of the  
XX same genes in one or more second filamentous fungal cells. Monitoring  
XX the global expression of genes from FF cells allows the production  
XX potential of the microorganisms to be improved. New genes may be  
XX discovered, possible functions of unknown open reading frames can be  
XX identified and gene copy number variation and stability can be  
XX monitored. The expression of genes can be used to study how FF cells  
XX adapt to changes in culture conditions, environmental stress, spore  
XX morphogenesis, recombination, metabolic or catabolic pathway  
XX engineering. Using ESTs provides several advantages over genomic or  
XX random cDNA clones including elimination of redundancy as one spot on an  
XX array equals one gene or open reading frame, and organization of the  
XX microarrays based on function of the gene products to facilitate  
XX analysis of the results. AAF07478 to AAF11247 represents ESTs from  
XX *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*  
XX *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and  
XX AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are  
XX all specifically claimed in the present invention.  
XX  
XX Sequence 1202 BP; 282 A; 294 C; 282 G; 344 T; 0 other;

Alignment Scores:  
Pred. No.: 1,28e-65 Length: 1202  
Score: 657.50 Matches: 138  
Percent Similarity: 72.82% Conservative: 12  
Best Local Similarity: 66.99% Mismatches: 19  
Query Match: 74.21% Indels: 37  
DB: 21 Gaps: 3

US-09-820-003b-2 (1-173) x AAFI4053 (1-1202)

OY 2 SerSerMetAspProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGly 21  
DB 111  
AC

Db 180 TCATCATGAACCTGAATACACTATCTCTCAAGCTCTTATCGAGATCCGGT 239  
 Qy 22 ValGlySerCysLeuLeuArpPheAlaAspThrTyThrGluSerTyrlle 41  
 Db 240 GTTGAATAATCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 299  
 Qy 42 SerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61  
 Db 300 TCCACTATGTTGTTGATTAAATCCGAAATCGAATCTGATGCAACAGTGAAG 359  
 Qy 62 LeuGlnIle----- 64  
 Db 360 CTTACGATTTGGACACTGCGCGCCAGAGCGGTCCGACCAATCAGCTCTTACTAT 419  
 Qy 65 -----GluSerPheAsnAsn 69  
 Db 420 CGAGGTGCTATGATGTGTGTGTATGATGTTACTGATGATGATGATGATGATGAT 479  
 Qy 70 ValIlyGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89  
 Db 480 GTGAAGCGTGGCTCCAGAGATGATGATGATGATGATGATGATGATGATGATGAT 539  
 Qy 90 ValGlyAsnLysCysAspLeuThrThrLysValValAspTyrThrThrAlaLysGlu 109  
 Db 540 GTGGGTACCAAGAGTACATGCAATGATGATGATGATGATGATGATGATGATGATGAT 599  
 Qy 110 PheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129  
 Db 600 TTCGCTGATGCTCTGGAATACATCTCTGAGACCTCTGTAAGAAATGCTGACAGCTC 659  
 Qy 130 GluIleSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr 149  
 Db 660 GAGAAAGCTTCTGCAATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 710  
 Qy 150 AlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerIly 169  
 Db 711 GCCACTGTCACACACACACACACACACACACACACACACACACACACACACACAC 770  
 Qy 170 -----GlyGlyCysCys 173  
 Db 771 TCCGAGGTGCTGCTCTGC 788  
 RESULT 10  
 AAC33987  
 ID AAC33987 standard; DNA: 959 BP.  
 AAC33987;  
 17-OCT-2000 (first entry)  
 Arabidopsis thaliana DNA fragment SEQ ID NO: 5039.  
 Hybridisation assay; genetic mapping; gene expression control;  
 protein identification; signal transduction pathway;  
 metabolic pathway; promoter; termination sequence; ss.  
 Arabidopsis thaliana.  
 EP1033405-A2.  
 06-SEP-2000.  
 25-FEB-2000; 2000EP-0301439.  
 25-FEB-1999; 99US-0121825.  
 05-MAR-1999; 99US-0123160.  
 09-MAR-1999; 99US-0123548.  
 23-MAR-1999; 99US-0125788.  
 23-MAR-1999; 99US-0126264.  
 29-MAR-1999; 99US-0126785.  
 01-APR-1999; 99US-0127462.  
 06-APR-1999; 99US-0128234.  
 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130044.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137502.  
 PR 04-JUN-1999; 99US-0137508.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140693.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.





DB 714 GCAGAGATCTAAGCCACCAACGATCCGAGCAGCAACCCGTTAACACGCA--- 770  
Oy 170 GLYGLYCYS 173  
DB 771 TCAGCTCTCTCC 782  
RESULT 11  
AAC35200  
ID AAC35200 standard; DNA; 932 BP.  
XX AAC35200;  
AC  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9347.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140358.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145917.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145915.  
PR 28-JUL-1999; 99US-0145916.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147199.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 9905-0148565.  
 PR 13-AUG-1999; 9905-0148684.  
 PR 16-AUG-1999; 9905-0149368.  
 PR 17-AUG-1999; 9905-0149175.  
 PR 18-AUG-1999; 9905-0149426.  
 PR 20-AUG-1999; 9905-0149722.  
 PR 20-AUG-1999; 9905-0149722.  
 PR 20-AUG-1999; 9905-0149722.  
 PR 20-AUG-1999; 9905-0149929.  
 PR 23-AUG-1999; 9905-0149930.  
 PR 23-AUG-1999; 9905-0149930.  
 PR 25-AUG-1999; 9905-0150566.  
 PR 26-AUG-1999; 9905-0150864.  
 PR 27-AUG-1999; 9905-0151065.  
 PR 27-AUG-1999; 9905-0151065.  
 PR 27-AUG-1999; 9905-0151080.  
 PR 30-AUG-1999; 9905-0151303.  
 PR 31-AUG-1999; 9905-0151438.  
 PR 01-SEP-1999; 9905-0151930.  
 PR 07-SEP-1999; 9905-0152363.  
 PR 10-SEP-1999; 9905-0153070.  
 PR 13-SEP-1999; 9905-0153758.  
 PR 15-SEP-1999; 9905-0154018.  
 PR 16-SEP-1999; 9905-0154039.  
 PR 20-SEP-1999; 9905-0154779.  
 PR 22-SEP-1999; 9905-0155139.  
 PR 23-SEP-1999; 9905-0155486.  
 PR 24-SEP-1999; 9905-0155659.  
 PR 28-SEP-1999; 9905-0156458.  
 PR 29-SEP-1999; 9905-0156596.  
 PR 04-OCT-1999; 9905-0157117.  
 PR 05-OCT-1999; 9905-0157793.  
 PR 06-OCT-1999; 9905-0157865.  
 PR 07-OCT-1999; 9905-0158029.  
 PR 08-OCT-1999; 9905-0158232.  
 PR 12-OCT-1999; 9905-0158369.  
 PR 13-OCT-1999; 9905-0159293.  
 PR 13-OCT-1999; 9905-0159293.  
 PR 13-OCT-1999; 9905-0159295.  
 PR 14-OCT-1999; 9905-0159329.  
 PR 14-OCT-1999; 9905-0159330.  
 PR 14-OCT-1999; 9905-0159331.  
 PR 14-OCT-1999; 9905-0159637.  
 PR 14-OCT-1999; 9905-0159638.  
 PR 18-OCT-1999; 9905-0159584.  
 PR 21-OCT-1999; 9905-0160741.  
 PR 21-OCT-1999; 9905-0160741.  
 PR 21-OCT-1999; 9905-0160767.  
 PR 21-OCT-1999; 9905-0160768.  
 PR 21-OCT-1999; 9905-0160770.  
 PR 21-OCT-1999; 9905-0160814.  
 PR 21-OCT-1999; 9905-0160815.  
 PR 22-OCT-1999; 9905-0160980.  
 PR 22-OCT-1999; 9905-0160981.  
 PR 22-OCT-1999; 9905-0160981.  
 PR 22-OCT-1999; 9905-0160981.  
 PR 25-OCT-1999; 9905-0161404.  
 PR 25-OCT-1999; 9905-0161405.  
 PR 25-OCT-1999; 9905-0161406.  
 PR 26-OCT-1999; 9905-0161359.  
 PR 26-OCT-1999; 9905-0161360.  
 PR 26-OCT-1999; 9905-0161361.  
 PR 28-OCT-1999; 9905-0161920.  
 PR 28-OCT-1999; 9905-0161922.  
 PR 28-OCT-1999; 9905-0161993.  
 PR 29-OCT-1999; 9905-0162142.

Alignment Scores:  
 Pred. No.: 5 83e-63  
 Score: 633.00  
 Percent Similarity: 70.798  
 Best Local Similarity: 65.848  
 Query Match: 71.448  
 DB: 21  
 Matches: 932  
 Conserved: 133  
 Mismatches: 10  
 Indels: 25  
 Gaps: 34

US-09-820-003b-2 (1-173) x AAC35200 (1-932)

OY 4 MetaspProglutyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGlyValGly 23  
 DB 125 ATGATACCTGCAATATGACTATACCTGCTTCAAGCTTCGCTTATTCGATTCGCTTGA 184  
 OY 24 LysSerGlyLeuLeuLeuGlyPheAlaAspPheThrThrGluSerThr 43  
 DB 185 AATCATGCTGCTTCAATGATTTGCGATGATTTCTACTGATGCTACATGACCC 244  
 OY 44 IleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGln 63  
 DB 245 ATTGGTGTGACTTTAAATTCGACAGTTGACGAGACGAAAGCATCAAACTCCAG 304  
 OY 64 Ile----- 64  
 DB 305 ATCTGGGACACAGCAGCCAGAACGTTTCAGACAACTACAGCTACTACAGAGA 364  
 OY 65 -----GluSerPheAsnValLys 71  
 DB 365 GCTCATGGATCATTTGCTACTTATGATGTCACAGACCTAGAGAGCTTCAACAGTCAAA 424  
 OY 72 GlnThrPheGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuValGly 91  
 DB 425 CATGCGCTGAATGAATATGACCGTACGCTACGCAAAATGTGACAAAGCTTCTGTTG 484  
 OY 92 AsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGluPheAla 111  
 DB 485 AACAGAACGATCTCACTTCAACAGAAAGTTGATCCACTGACGACAGCTTACGCTTGGCA 544  
 OY 112 AspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln 131  
 DB 545 GATTAACCTTGATCCATCTTCTTGAAACAAAGCTGTAAATCAACAAAGTGAAGAA 604  
 OY 133 SerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThrAlaGly 151  
 DB 605 GCTTTCATGGCTATGACTGCTGCAATTAGAACAAAGATGCT---AGCAACACGCTGGA 661  
 OY 155 GlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysIleSerGlyGly 171  
 DB 662 GGTCCAAAGCAGCAACGATCCAGATCCGTCGACAGCAGCAGTGAACAGCAAA---TCAGGC 718  
 OY 172 CysCys 173  
 DB 719 TGTGT 724  
 RESULT 12  
 ID AAS83860/c  
 AC AAS83860; standard; cDNA; 1193 BP.  
 AC AAS83860;  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #19664.  
 DE Human: Chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 PN W0200175067-A2.  
 PD 11-OCT-2001.  
 PP 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PR (HYSE-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI

DR WPI: 2001-639362/73.  
DR P-PSDB: ABG19673.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 1; SEQ ID No 19664; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1193 BP; 372 A; 279 C; 213 G; 329 T; 0 other;

Alignment Scores:

Pred. No.: 8.64e-62 Length: 1193  
Score: 624.00 Matches: 129  
Percent Similarity: 74.57% Conservative: 0  
Best Local Similarity: 74.57% Mismatches: 0  
Query Match: 70.43% Indels: 44  
DB: Gaps: 1

US-09-820-003b-2 (1-173) x AAS83860 (1-1193)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
DB 1003 ATGTCACAGCATGATCCGATATGATTAATTCAGTACTGATTCGACACATCA 944  
QY 21 GlyAlaGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
DB 943 GGGGTGGAAAGCTCTGCTCTTCTAGGTTTGCAGATGATACATATACAGAAAGCTAC 884  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 883 ATCAGCAACATTTGGTGTGATTCATAAATAGAACTATAGATGACGGGAAAAATATC 824  
QY 61 LysLeuGlnIleGluSerPheAsnAsnValLysGlnTrrLeuGlnIleLaspArgTyr 80  
DB 823 AACCTCAATAA----- 812  
QY 81 AlaSerGluAsnValAsnLysLeuLeuValGlyAsnLysCysAspLeuThrThrLysLys 100  
DB 812 ----- 812  
QY 101 ValValAspTyrThrThrAlaLysGluPheAlaAspSerLeuGlyIleProPheLeuGlu 120  
DB 811 -----GAATTTCTGATTCCTTCCTTGGATTCCTTTTGGAA 776  
QY 121 ThrSerAlaLysAsnAlaThrAsnValGluGlnSerPheMetThrMetAlaAlaGluIle 140  
DB 775 ACCAGTGCTAGATGACAGCAATGTAAGACAGTCTTCATGACAGATGGCAGCTGAGATT 716  
QY 141 LysLysArgMetGlyProGlyAlaThrAlaGlyLysGluLysSerAsnValLysIle 160

DB 715 AAAAAAGCAATGGGCTCCGAGACAAACACTGTGTGTGAGAAATGTTAAAT 656  
QY 161 GlnSerThrProValLysGlnSerGlyGlyLysCys 173  
DB 655 CAGACACTCCAGTCACAGCACTGAGTGAGATTGCTTC 617

RESULT 13

ID AAC42684  
AAC42684 standard; DNA; 777 BP.

AC AAC42684;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 36471.

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132861.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 18-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 21-JUN-1999; 99US-0139763.  
 PR 22-JUN-1999; 99US-0139817.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 02-JUL-1999; 99US-0142154.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143628.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147205.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 15-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149724.  
 PR 22-AUG-1999; 99US-0149920.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 26-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154772.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156596.  
 PR 29-SEP-1999; 99US-0156598.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 18-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
 Pred. No.: 1.61e-61  
 Score: 619.50  
 Percent Similarity: 68.97%  
 Best Local Similarity: 61.58%  
 Query Match: 69.92%  
 DB: 21  
 US-09-820-003b-2 (1-173) x AAC42684 (1-777)  
 Length: 777  
 Matches: 125  
 Conservative: 15  
 Mismatches: 30  
 Indels: 33  
 Gaps: 2

```

OY      3 SerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuLleGlyAspSerGlyVal 22
DB      163 ACCATGAAATCGTAGACATATCTTTCAAGCTCGCTATCGGGATCTGGCGTA 222
OY      23 GlyLysSerCysLeuLeuLeuArgPheAlaAspPheThrTyrThrGlnSerYrLleSer 42
DB      223 GGGCAATCTTCTCTTCTTTGAGATCTCTGATGATCTTATGTAAGATTACATTTAGC 282
OY      43 ThrIleGlyValAspPheLysLleArgThrIleGluLeuAspGlyLysThrIleLysLeu 62
DB      283 ACTATTGGAGTCGATTTTAAATTTAGACTGTGAAACAAGATGGCAAAACATTTAAGCTC 342
OY      63 GlnIle----- 64
DB      343 CAATTTGGGACACTGCTGCTCAGACGCTTACAGACTATTACTAGACGTTACTACCGT 402
OY      65 -----GlnSerPheAsnVal 70
DB      403 GGGGACATGCAATTATTATTGCTACAGATGTCACAGATGAAGAAAGCTCAATATATGTC 462
OY      71 LysGlnTrpLeuGlnGluLleAspArgTyrAlaSerGluAsnValAsnLysLeuVal 90
DB      463 AAGCAATGTTGAGTGAATGATGCTATGCTACTGACAAATGTCACAAACTCTTGT 522
OY      91 GlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGluPhe 110
DB      523 GGAACAAGTGTGATCTACTGTAAGAAACAGAGCCATCTCTTATGAAACGCCAAGCTTTT 582
OY      111 AlaAspSerLleGluLleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnValGlu 130
DB      583 GCCCATGAAATCGGATTCCTTTATGAGACTACTGCAAAAGATGCTACAAACTGAGTA 642
OY      131 GlnSerPheMetThrMetAlaAlaGluLleLysLysArgMetGlyProGlyValAlaThrAla 150
DB      643 CAGGCTTCATGAGCAATGCTGCATCCAAAGAGACAAAGGCTAGCAACACCGCTGG 702
OY      151 GlyGlyAlaGluLysSerAsnValLysLleGlnSerThrProValLysGlnSerGly 170
DB      703 AATAATGCAAGACCAACCGATCGATGATGAGAGACAGACGCTGTGACAGAGAG--AAC 759
OY      171 GlyCysCys 173
DB      760 GGCCTGCTGC 768

```

RESULT 14  
ABN98761 standard; DNA; 881 BP.

ABN98761;  
01-AUG-2002 (first entry)  
Arabidopsis thaliana expressed polynucleotide SEQ ID NO 529.  
Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;  
disease; crop; thale cress; tolerance factor; insect; pathogen;  
nutrition; ds.  
Arabidopsis thaliana.  
US2002023281-A1.  
21-FEB-2002.  
26-JAN-2001; 2001US-0770445.  
27-JAN-2000; 2000US-178472P.

PA (RAIN/) RAINES T M.  
PA (YUY/) YU Y.  
PA (RAME/) RAMEKA J G.  
PA (PAGE/) PAGE A.  
PA (MATH/) MATHEN A V.  
PA (LEDF/) LEDFORD B L.  
PA (WOES/) WOESSNER J P.  
PA (HAAS/) HAAS W D.  
PA (GARC/) GARCIA C A.  
PA (KRICK/) KRICKER M.  
PA (SLAT/) SLATER T.  
PA (DAVI/) DAVIS K R.  
PA (ALE/) ALLEN K.  
PA (HOFF/) HOFFMAN N.  
PA (HURB/) HURBAN P.

PI Gorlach J, An Y, Hamilton CM, Price JJ, Raines TM, Yu Y;  
PI Ramakea JC, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;  
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;  
PI Hurban P;  
PI WPI: 2002-400781/43.

DR New Arabidopsis thaliana nucleic acid for identifying homologous genes,  
XX producing compositions that modulate the expression or function of its  
XX encoded protein, and mapping functional regions of protein

PT Claim 1; SEQ ID NO 529; 49pp + Sequence Listing; English.

PS The invention relates to an Arabidopsis thaliana nucleic acid (I)  
XX comprising a sequence capable of hybridizing under stringent conditions  
XX to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),  
XX given in the specification or its fragment. A polypeptide (II) encoded by  
XX (I), a transgenic plant (III) comprising an exogenous nucleic acid or a  
XX genetically modified cell (IV) comprising an exogenous nucleic acid, is  
XX useful for screening a candidate agent for its biological effect. (I) is  
XX useful in identifying homologous or related genes, in producing  
XX compositions that modulate the expression or function of its encoded  
XX protein, mapping functional regions of the protein and in studying  
XX associated physiological pathways. (I) is also useful for the genetic  
XX manipulation of cells, particularly plant cells. (I) is also useful in  
XX screening assays of various plant strains to determine the strains that  
XX are best capable of withstanding a particular disease or environmental  
XX stress. (II) and (III) are useful for screening of biologically active  
XX agents, e.g. fungicides, insecticides, etc., for elucidating biochemical  
XX pathways. The screened agents are useful in improved methods of treating  
XX crops to prevent or treat disease. (II) are also useful in screening  
XX programs to identify agents that mimic or enhance the action of tolerance  
XX factors. Such agents are useful in improved methods of treating crops to  
XX enhance their tolerance to environmental stress. (I) is also useful  
XX for enhancing or inhibiting production of a biosynthetic product in a  
XX plant. (III) is useful for identifying other mediators that may induce  
XX expression of proteins of interest, for establishing the extent to which  
XX any specific insect and/or pathogen is responsible for damage to a  
XX particular plant, for identifying other mediators that enhance or induce  
XX tolerance to environmental stress, for identifying factors involved in  
XX biosynthetic pathways of nutritional, commercial, or medicinal value and  
XX for identifying productions of nutritional, commercial or medicinal  
XX value. (IV) is useful in the study of genetic function and regulation,  
XX for alteration of the cellular metabolism and for screening compounds  
XX that may affect the biological function of the gene or gene products.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from USPTO  
XX at seqdata.uspto.gov/sequence.html?DocID=999909770445.

SO Sequence 881 BP; 241 A; 199 C; 180 G; 261 T; 0 other.

Alignment Scores:  
Pred. No.: 1.9e-61 Length: 881  
Score: 619.50 Matches: 125  
Percent Similarity: 68.97% Conservative: 15  
Best Local Similarity: 61.58% Mismatches: 30  
Query Match: 69.92% Indels: 33

[illegible]

```

PR 16-JUL-1999; 99US-0144085;
PR 16-JUL-1999; 99US-0144086;
PR 19-JUL-1999; 99US-0144325;
PR 19-JUL-1999; 99US-0144326;
PR 19-JUL-1999; 99US-0144331;
PR 19-JUL-1999; 99US-0144332;
PR 19-JUL-1999; 99US-0144333;
PR 19-JUL-1999; 99US-0144334;
PR 19-JUL-1999; 99US-0144335;
PR 20-JUL-1999; 99US-0144336;
PR 20-JUL-1999; 99US-0144632;
PR 20-JUL-1999; 99US-0144633;
PR 21-JUL-1999; 99US-0144814;
PR 21-JUL-1999; 99US-0145086;
PR 21-JUL-1999; 99US-0145087;
PR 22-JUL-1999; 99US-0145088;
PR 22-JUL-1999; 99US-0145089;
PR 22-JUL-1999; 99US-0145090;
PR 22-JUL-1999; 99US-0145192;
PR 23-JUL-1999; 99US-0145193;
PR 23-JUL-1999; 99US-0145218;
PR 23-JUL-1999; 99US-0145219;
PR 26-JUL-1999; 99US-0145224;
PR 26-JUL-1999; 99US-0145276;
PR 27-JUL-1999; 99US-0145913;
PR 27-JUL-1999; 99US-0145918;
PR 27-JUL-1999; 99US-0145919;
PR 28-JUL-1999; 99US-0145951;
PR 02-AUG-1999; 99US-0146386;
PR 02-AUG-1999; 99US-0146387;
PR 02-AUG-1999; 99US-0146388;
PR 02-AUG-1999; 99US-0146389;
PR 03-AUG-1999; 99US-0147038;
PR 04-AUG-1999; 99US-0147204;
PR 04-AUG-1999; 99US-0147302;
PR 05-AUG-1999; 99US-0147192;
PR 05-AUG-1999; 99US-0147260;
PR 06-AUG-1999; 99US-0147303;
PR 06-AUG-1999; 99US-0147416;
PR 09-AUG-1999; 99US-0147493;
PR 09-AUG-1999; 99US-0147935;
PR 10-AUG-1999; 99US-0148171;
PR 11-AUG-1999; 99US-0148319;
PR 12-AUG-1999; 99US-0148341;
PR 13-AUG-1999; 99US-0148565;
PR 13-AUG-1999; 99US-0148684;
PR 16-AUG-1999; 99US-0149368;
PR 17-AUG-1999; 99US-0149175;
PR 18-AUG-1999; 99US-0149426;
PR 20-AUG-1999; 99US-0149722;
PR 20-AUG-1999; 99US-0149723;
PR 20-AUG-1999; 99US-0149929;
PR 23-AUG-1999; 99US-0149902;
PR 23-AUG-1999; 99US-0149930;
PR 25-AUG-1999; 99US-0150566;
PR 26-AUG-1999; 99US-0150884;
PR 27-AUG-1999; 99US-0151065;
PR 27-AUG-1999; 99US-0151066;
PR 27-AUG-1999; 99US-0151080;
PR 30-AUG-1999; 99US-0151303;
PR 31-AUG-1999; 99US-0151438;
PR 01-SEP-1999; 99US-0151930;
PR 07-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 13-SEP-1999; 99US-0154018;
PR 16-SEP-1999; 99US-0154039;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155653;
PR 28-SEP-1999; 99US-0156458;
PR 29-SEP-1999; 99US-0156596;
PR 04-OCT-1999; 99US-0157117;
PR 05-OCT-1999; 99US-0157753;
PR 06-OCT-1999; 99US-0157865;

```

```

PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158325;
PR 13-OCT-1999; 99US-0158293;
PR 13-OCT-1999; 99US-0158294;
PR 13-OCT-1999; 99US-0158295;
PR 14-OCT-1999; 99US-0158329;
PR 14-OCT-1999; 99US-0158330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159337;
PR 18-OCT-1999; 99US-0159638;
PR 18-OCT-1999; 99US-0159584;
PR 21-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 21-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161922;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;

```

## Alignment Scores:

```

Pred. No.: 9,5e-61      Length: 666
Score: 612.00          Matches: 128
Percent Similarity: 70.56%      Conservative: 11
Best Local Similarity: 64.97%    Mismatches: 24
Query Match: 69.07%             Indels: 34
DB: 21                      Gaps: 3

```

US-09-820-003b-2 (1-173) x AAC42764 (1-666)

```

QY 9 Asprytleupherylsleuleuileglyaspserglyalglyssercysleuenu 28
    |||||||
DB 73 GACTATTGTCACAGCTTTCCTCATGCTGATTCGTGGAAAGTCGTTGCTT 132
    |||||||

QY 29 LeuArgPheAlaAspThrThrThrGluSerThrIleValAspPhe 48
    |||||||
DB 133 CTAAAGTTGCTGATGATTCCTACCTGATGATGATGATGATGATGATGATGAT 192
    |||||||

QY 49 LysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGluIle----- 64
    |||||||
DB 193 AAAATCCGACAGTCGACAGATGAGAAAGCAATCCAGATTTGGATACGCA 252
    |||||||

QY 64 ----- 64
DB 253 GGCCAGAACGCTTCAGAACGATTAAGTACTACAGAGAGGCTCATGATATT 312
    |||||||

QY 65 -----GluSerPheAsnAsnValLysGlnTrpLeuGluGlu 76
    |||||||
DB 313 CTGACTTATGATGTACAGATCTAGAAAGCTTTCACAGCAAGCAATGGCTAAAGAA 372
    |||||||

QY 77 IleAspArgTyrAlaSerGluAsnValAsnLysLeuValGlyAsnLysCysAspLeu 96
    |||||||
DB 373 ATGACCGCTATGACAGAGGAAAGTTTACAGAGTACGTTGGGAACAAGTGTGATCTC 432
    |||||||

QY 97 ThrThrLysLysValValAspTyrThrThrAlaLysGluPheAlaAspSerLeuGlyIle 116
    |||||||
DB 433 ACATCACAGAAAGTGTATCCACTGACAGACACTAAGGCTTCGCTGATGATGATGATG 492
    |||||||

QY 117 ProPheLeuGluTrpSerAlaLysAsnAlaThrAsnValGluGlnSerPheMetThrMet 136
    |||||||
DB 493 CCATTCTTGAAACAGAGCTTAAGAAATGCTACCAATGCTGAAGAGCTTTCATGGCCATG 552
    |||||||

```



Oy 137 AlaAlaGluIleLysLysArgMetGlyProGlyAlaThrAlaGlyGlyAlaGlyLysSer 156  
DB 553 ACTGCTGCATTCAGACGACGAAATGCCG--AGCCAACTGCGAGGAGATCTAAGCCACCA 609  
Oy 157 AsnValLysTlleGlnSerThrProValLysGlnSerGlyGlyLysCys 173  
DB 610 ACGGTCCAGATCCGAGGACAACTGTATTACCAACAA---TCAGGCTGCTGC 657

Search completed: June 22, 2003, 18:04:34  
Job time : 221 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 22, 2003, 16:28:49 ; Search time 1618 Seconds

(without alignments)  
311.734 Million cell updates/sec

Title: US-09-820-003b-2

Perfect score: 886  
Sequence: 1 MSSMNPEDYDFKLLIGDS.....EKSNAKIQSTPVKSGGGCC 173

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame\_plus.model -DRV-xlp  
-Q/cgn2.1/USPFO.spool/US09820003/funat.18062003.144129.13092/app-query.fasta.1.327  
-DB-GenEmbl -QEM-fastap -SUFPI-p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45  
-DOCALLGN=200 -THR.SCORE-pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pic -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09820003.ecgn.1.1.3745-etunat.18062003.144129.13092 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEDUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONOLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vl.\*  
15: gb.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.ov.\*  
22: em.pat.\*  
23: em.ph.\*  
24: em.pl.\*  
25: em.ro.\*  
26: em.sts.\*  
27: em.un.\*  
28: em.un.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860	97.1	615	6 E03861	E03861 cDNA encod
2	860	97.1	618	9 AF498929	AF498929 Homo sapi
3	860	97.1	723	9 HUMRAB1A	M28209 Homo sapien
4	860	97.1	840	4 CRRAB1	X56384 Canine RAD1
5	860	97.1	1428	10 MAFPT1	I00094 Mouse mRNA
6	860	97.1	1433	9 BC000905	BC000905 Homo sapi
7	860	97.1	1444	10 BC002077	BC002077 Mus muscu
8	860	97.1	2686	10 AF226873	AF226873 Mus muscu
9	849	95.8	618	10 NATRASA	J02998 Rat ras-rel
10	838	94.6	841	5 DYGORAB1	M38393 Discopyle o
11	768.5	86.7	925	6 AR070365	AR070365 Sequence
12	768.5	86.7	1985	6 AX086155	AX086155 Sequence
13	768.5	86.7	1985	9 HSM801608	AL136635 Homo sapi
14	762.5	86.1	654	10 RNRAB1B	X13905 Rat cdna fo
15	762.5	86.1	161955	2 AC106368	AC106368 Mus muscu
16	761.5	85.9	1859	10 BC016408	BC016408 Mus muscu
17	758.5	85.6	151988	2 AC123143	AC123143 Rattus no
18	740.5	83.6	1858	9 HSA245875	AJ245875 Homo sapi
19	737.5	83.2	185285	2 AC094411	AC094411 Rattus no
20	733.5	82.8	169095	2 AL139414	AL139414 Homo sapi
21	733.5	82.8	194976	9 AL513165	AL513165 Human DNA
22	733.5	82.8	207945	2 AC117841	AC117841 Rattus no
23	733.5	82.7	718	3 LSRAB1	X72688 L. stagnalis
24	722	81.5	1086	10 RNU63023	U63023 Rattus nov
25	722	81.5	174028	2 AC099354	AC099354 Rattus no
26	700	79.0	2257	9 AK055927	AK055927 Homo sapi
27	683	77.1	1772	3 DB4312	DB4312 Drosophila
28	671	75.7	952	3 AF013572	AF013572 Bombyx mo
29	658.5	74.3	720	8 CNS019BR	AL111439 Botrytis
30	655	73.9	1868	8 AY121695	AY121695 Drosophila
31	645	72.8	979	8 LJRAB1C	Z73932 L. japonicus
32	644	72.7	938	8 AY052204	AY052204 Arabidops
33	644	72.7	950	8 AF127134	AF127134 Arabidops
34	636.5	71.8	864	8 PEXGTPBP10	D12549 Pea mRNA fo
35	634.5	71.6	609	8 AY060495	AY060495 Arabidops
36	633	71.4	636	8 CNS01BM9	AL114769 Botrytis
37	633	71.4	640	8 AY114013	AY114013 Arabidops
38	633	71.4	931	8 AF324990	AF324990 Arabidops
39	633	71.4	932	8 AY085384	AY085384 Arabidops
40	633	71.4	939	8 AY080753	AY080753 Arabidops
41	632.5	71.4	1008	8 NPA33CTP	Y08425 N. plumbagin
42	632	71.3	1074	8 AB024994	AB024994 Cicer atri
43	630.5	71.2	1129	8 LJRAB1D	Z73933 L. japonicus
44	628.5	70.9	898	8 NTRKBP	X72212 L. tadacum m
45	626.5	70.7	860	8 LJRAB1E	Z73934 L. japonicus

## ALIGNMENTS

RESULT 1

E03861 LOCUS E03861 615 bp RNA linear PAT 29-SEP-1997  
 DEFINITION cDNA encoding GTP binding protein yptl.  
 ACCESSION E03861  
 VERSION E03861.1 GI:2172075  
 KEYWORDS JP 1992226000-A/3.  
 SOURCE Mus sp.  
 ORGANISM Mus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 615)  
 Yamauchi K., Uchida K., Takai Y., Aisaka K. and Sakurada K.  
 LOW MOLECULAR WEIGHT PROTEIN BINDING TO GUANOSINE TRIPHOSPHATE  
 Patent: JP 1992226000-A 3 14-AUG-1992;  
 KYOMA HAKKO KOGYO CO LTD  
 COMMENT OS Mus sp. (mouse)  
 PN JP 1992226000-A/3  
 PD 14-AUG-1992  
 PF 25-APR-1991 JP 1991095304  
 PR 21-MAY-1990 JP 90P 130547  
 PI YAMAGUCHI KAZUO, UCHIDA KAZUHISA, TAKAI YOSHIMI, AISAKA KAZUO,  
 PI SAKURADA KAZUHIRO  
 PC C07K15/18,C07K13/00,C07K15/06,C12N15/12,G01N33/50//A61K37/02;  
 CC topology: linear.  
 CC strandedness: single;  
 CC location:Qualifiers  
 FEATURES  
 source 1..615  
 /organism="Mus sp."  
 /db\_xref="taxon:10095"  
 BASE COUNT 196 a 117 c 153 g 149 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,51e-75 Length: 615  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservatve: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 DB: Gaps: 1  
 US-09-820-003b-2 (1-173) x E03861 (1-615)  
 QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuLeuIleGlyAspSer 20  
 DB 1 ATGTCACAGATGAATCCGCAATATGATTTATTCAGTACTTCTGATGGCGATTCCT 60  
 QY 21 GlyValIGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 DB 61 GGGGTGGAAATCCTGCTTCTCTCTAGAGTTGCGAGATGATACGATACGGAAGCTTAC 120  
 QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyTyrThrIle 60  
 DB 121 ATTCACACAAATGCTGTGATTTCAAGATACGAACTATGAGTTAGTGGAAACCAATTC 180  
 QY 61 LysLeuGlnIle----- 64  
 DB 181 AAGCTACAGATATGGGACACAGCAGCCAGGAAGATTTGCAACATCTTCAGTTAT 240  
 QY 65 -----GluSerPheasn 68  
 DB 241 TACAGAGAGCCCATGGCATCATAGTTGTGTATGATGTGACAGATCAGAGATGCTTCAT 300  
 QY 69 AsnValLysGlnTrrPheGlnGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
 DB 301 AACGTTAAACAGTGTGCTGACGAGATAGATCCCTACGCGAGGAAATGTCACAAAGTTG 360  
 QY 89 LeuValIGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
 DB 361 TTGGTAGGGAACAATGTGACTGACACACAAAGAAAGTAGTACTACACAAACAGCAAG 420  
 QY 109 GluPheAlaAspSerLeuGlyTyrLeuPheLeuGlnIleThrSerAlaLysAsnAlaThrAsn 128  
 DB 421 GAATTTGCAGATTCCTGCTTCTCTAGAGTTGCGAGATGATACATATACAGAAAGCTTAC 480

QY 129 ValGluGlnSerPheMetThrMetAlaAlaGlnIleLysLysArgMetGlyProGlyAla 148  
 DB 481 GTAGAACACTCTTTCATGACGATGACGAGCTGAGATTAATAACGCAATGGCTCTGAGCT 540  
 QY 149 ThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 DB 541 ACAGCTGCTGTGGCCGAGAAATGTCATTTAAATCCAGACACTCCAGTCAAGCAGTCA 600  
 QY 169 GlyGlyLysCysCys 173  
 DB 601 GGTGAGGCTGCTGC 615  
 RESULT 2  
 AF498929 618 bp mRNA linear PRI 01-MAY-2002  
 LOCUS AF498929  
 DEFINITION Homo sapiens small GTP binding protein RAB1A (RAB1A) mRNA, complete  
 cds.  
 ACCESSION AF498929  
 VERSION AF498929.1 GI:20379033  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 618)  
 Puhl,H.L. III, Ikeda,S.R. and Aronstam,R.S.  
 Homo sapiens RAB family small GTP binding protein RAB1A  
 Unpublished  
 2 (bases 1 to 618)  
 Puhl,H.L. III, Ikeda,S.R. and Aronstam,R.S.  
 Direct Submission  
 Submitted (05-APR-2002) CDNA Resource Center, Guthrie Research  
 Institute, One Guthrie Square, Sayre, PA 18840, USA  
 FEATURES  
 source 1..618  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="Brain"  
 1..618  
 /gene="RAB1A"  
 /gene="RAB1A"  
 /note="RAB family member"  
 /codon\_start=1  
 /product="small GTP binding protein RAB1A"  
 /protein\_id="AA021077.1"  
 /db\_xref="GI:20379034"  
 /translation="MSSMNPEDYDFKLLITDSCVGSCLLRADPTYESYIST  
 GVDKIRITLEDGKTKIKLQIMDTAQDERRTTSYRGARGLIYVYVDTQSENNY  
 KQMLQIDRYASENVNKLIVGKKCDLTKRKVDYTAKEFAADSLGIPLETSAKNATN  
 VQSEFTMAAEIKKRMGPATFGAEKSVKIOSIPVQSGGCG"  
 BASE COUNT 203 a 110 c 149 g 156 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,53e-75 Length: 618  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservatve: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 DB: Gaps: 1  
 US-09-820-003b-2 (1-173) x AF498929 (1-618)  
 QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuLeuIleGlyAspSer 20  
 DB 1 ATGTCACAGATGAATCCGCAATATGATTTATTCAGTACTTCTGATGGCGACCTCA 60  
 QY 21 GlyValIGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 DB 61 GGGGTGGAAATCCTGCTTCTCTAGAGTTGCGAGATGATACGATACGGAAGCTTAC 120

41 IISerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB ATCAGACCAATGTCGTGATTTCCAAATAGACATATAGAGTTAGACGGGAAAAACATC 180  
QY LysLeuGlnIle----- 64  
DB 181 AGCTTCAAATATGAGACACAGACGCCAGAAAGATTTCACAAATCCTCCAGTTAT 240  
QY 65 -----GluSerPheAsn 68  
DB 241 TACAGAGACCCATGCGATCATGTTGTTGATATGACAGATCAGGAGTCTTCAT 300  
QY 69 AsnValLysGlnIleGluGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
DB 301 AATGTTAAACAGTGGCTGACGAAATAGATCGTTAGCCAGTGAATGCAACAAATTCG 360  
QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysValValAspTyrThrAlaLys 108  
DB 361 TTGAGAGGAAACAAATGATCTGACACAAAGAAAGTAGTACACTACAAACAGCAG 420  
QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
DB 421 GAATTCGCTGATCCCTTGGAATTCGCTTTTGGAAACCACTGCTAAGAAATCAAGAT 480  
QY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
DB 481 GTAGACACAGCTTTCATGACATGCGACGATGATTTAAAAAGCAATGGCTCCGAGACA 540  
QY 149 ThrAlaGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
DB 541 ACAGCTGGTGGTCTGACAGATCCAAATGTTAAATTCAGACACTCCAGTCAAGACGTCA 600  
QY 169 GlyGlyLysCysCys 173  
DB 601 GGTGGAGGTTGCTGC 615  
RESULT 3  
LOCUS HUMRABIA 723 bp mRNA linear PRI 08-JAN-1995  
DEFINITION Homo sapiens GTP-binding protein (RAB1) mRNA, complete cds.  
ACCESSION M28209.1 GI:550059  
VERSION M28209.1 GI:550059  
KEYWORDS GTP-binding protein; ras oncogene.  
SOURCE Homo sapiens (tissue library: of J.Mallet) pheochromocytoma cdna to mRNA.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 723)  
Zahraoui A, Touchot N., Chardin P. and Tavilian A.  
The human Rab genes encode a family of GTP-binding proteins related  
to yeast YPL1 and Sec4 products involved in secretion  
J Biol. Chem. 264 (21), 12394-12401 (1989)  
JOURNAL  
MEDLINE  
8930868  
2301306  
COMMENT  
On Sep 27, 1994 this sequence version replaced gi:341862.  
FEATURES  
Source  
1..723  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="Unassigned"  
/tissue\_type="pheochromocytoma"  
/tissue\_lib="Of J.Mallet"  
1..723  
/gene="RAB1"  
/gene="RAB1"  
51..668  
/codon\_start=1  
/product="GTP-binding protein"  
/protein\_id="AA60240.1"  
/db\_xref="GI:550060"  
/db\_xref="GB:G00-118-857"  
/translation="MSSMNPEDYDFLKLLIGDSGVKSCLLIRADDTYESTISTI  
gene  
CDS  
51..668  
/codon\_start=1  
/product="GTP-binding protein"  
/protein\_id="AA60240.1"  
/db\_xref="GI:550060"  
/db\_xref="GB:G00-118-857"  
/translation="MSSMNPEDYDFLKLLIGDSGVKSCLLIRADDTYESTISTI

GVDFKRTIIEIDGKTIKIQIMDTAGOEPRFTSSVYSGAGIIYVYDTPDESNVY  
KQWLEIDRVASENVKILYNGKCDITTKYVDYTAHERPDSGIGIFPLETSKNAIN  
VQDSMTAAEIKRBPGRATGCKSEKSNYKISTVAKSGGGCC"  
BASE COUNT 225 a 142 c 180 g 176 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,46e-75 Length: 723  
Score: 860.00 Matches: 173  
Percent Similarity: 84.39% Conservative: 0  
Best Local Similarity: 84.39% Mismatches: 0  
Query Match: 97.07% Indels: 32  
DB: 9 gaps: 1  
US-09-820-003b-2 (1-173) x HUMRABIA (1-723)  
QY 1 MetSerSerMetAsnProGluTyrAspTyrIleuPheLysLeuLeuIleGlyAspSer 20  
DB 51 ATGTCGACGATGAAATCCGAAATATGATTTATTCAAGTTACTTCGATTTGGGACTCA 110  
QY 21 GlyValGlyLysSerCysIleuLeuAspPheAlaAspSerThrTyrThrGluSerTyr 40  
DB 111 GGGTGGAAATCTCTGCTCTCTCTTACGTTTCACATATATACGAAAGCTAC 170  
QY 41 IISerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 171 ATCAGCACAAATGGTGGTGGATTTCAAAATAGAACTATAGAGTTAGACGGGAAACAAATC 230  
QY 61 LysLeuGlnIle----- 64  
DB 231 AAGCTTCAAATATGAGACACGACGACGCGAGAAAGATTTCGAACATCCTCCAGTTAT 290  
QY 65 -----GluSerPheAsn 68  
DB 291 TACAGAGACCCATGCGATCATGATGTTGATGATGATGACAGATCAGGAGTCTTCAT 350  
QY 69 AsnValLysGlnIleGluGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
DB 351 AATGTTAAACAGTGGCTGCGAGAAATAGATCGTATGCGCAATGCAAAATGTCACAAATTC 410  
QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
DB 411 TTGGTGGGAAACAAATGATGATCGACACAAAGAAATAGTACTACACACAGCAGAG 470  
QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
DB 471 GAATTCGCTGATCCCTTGGAATTCGCTTTTGGAAACGAGTCTAGAACAGACGACAC 530  
QY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
DB 531 GTAGACACAGCTTTCATGACGATGCGACGCTGAGATTAAGGAATGGGTCCCGGAGCA 590  
QY 149 ThrAlaGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
DB 591 ACAGCTGGTGGTGGTGGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAACAGTCA 650  
QY 169 GlyGlyLysCysCys 173  
DB 651 GGTGGAGGTTGCTGC 665  
RESULT 4  
LOCUS CEFAB1 840 bp mRNA linear MAM 30-SEP-1999  
DEFINITION Canine rab1 mRNA for ras-related GTP-binding protein involved in membrane traffic.  
ACCESSION X56384.1 GI:913  
VERSION X56384.1 GI:913  
KEYWORDS GTP-binding protein; protein transport; ras gene; ras gene family.  
SOURCE Canis familiaris.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
REFERENCE  
1 (bases 1 to 840)

AUTHORS Chavrier, P., Vingron, M., Sander, C., Simons, K. and Zerial, M.  
 TITLE Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell line  
 JOURNAL Mol. Cell. Biol. 10 (12), 6578-6585 (1990)  
 MEDLINE 91061765  
 PUBMED 2123294  
 REFERENCE 2 (bases 1 to 840)  
 AUTHORS Zerial, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-AUG-1990) Zerial M., EMBL, Meyerhofstrasse 1, 6900 Heidelberg, FRG

FEATURES  
 source Location/Qualifiers  
 1..840  
 /organism="Canis familiaris"  
 /strain="cockier spaniel"  
 /db\_xref="taxon:9615"  
 /clone="47"  
 /cell\_line="MDCK"  
 149..766  
 /gene="rab1"  
 149..766  
 /gene="rab1"  
 /function="GTP binding protein"  
 /codon\_start=1  
 /protein\_id="CAB56775.1"  
 /db\_xref="GI:6006435"  
 /translation="MSSNPEDYLFKLLIGDSGVKSCLLRFADPTYESYSTI  
 GVDKIRTELDGKTIKLIQIMPTAGOEPRRTTSSYRAGHIIYVVDQESNNV  
 KQWLEIDRYASENVNKLIVGKCDLTKKVVDTTAKAFACGPIPLETSARKATN  
 VQSPWTMAAEIKRMRGPRATAGAGKSNVAKIQSTPVKQSGGCC"

BASE COUNT 250 a 176 c 220 g 194 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6,52e-75 Length: 840  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 32  
 Query Match: 97.07% Indels: 32  
 DB: 4 Gaps: 1

US-09-820-003B-2 (1-173) x CFRAB1 (1-840)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 |||||  
 DB 149 ATGTCACGACGATGAATCCGAATATGATATTTATTCACACTTCTGATGGCGACTCT 208  
 |||||  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuAlaGlyPheAlaAspAspThrTyrThrGluSerTyr 40  
 |||||  
 DB 209 GGGGTTGGAAAGCTTGGCTCTCTTCTAGGTTGCAGATGATACATATACAGAAACCTAC 268  
 |||||  
 QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
 |||||  
 DB 269 ATCAGACACATGCTGTGATTTCCAAAATAGACATATAGAGTTAGATGGAAACAAATC 328  
 |||||  
 QY 61 LysLeuGlnIle----- 64  
 |||||  
 DB 329 AAGCTTCAAAATATGGACACAGCGCCCAAGAAAGATTTCGAACAATCAGCTCCAGTTAT 388  
 |||||  
 QY 65 -----GluSerPheAsn 68  
 |||||  
 DB 389 TACAGAGAGGCCATGGCATCATAGTGTGTATGACGTGACAGATCAGAGATCCCTCAAT 448  
 |||||  
 QY 69 AsnValLysGlnTyrLeuGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
 |||||  
 DB 449 AATGTTAAACATGTGGCTGACGAAATAGACCGTTATGCCAGTGAAGAAAGCTCAACAGTTG 508  
 |||||  
 QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValAlaAspTyrThrAlaLys 108  
 |||||  
 DB 509 TTGTAGAGGAGAACAAATGCGATCTGACCAAGAAAGATGATGACTACACAGCAAG 568  
 |||||  
 QY 109 GluPheAlaAspSerLeuGlyTyrProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
 |||||

DB 569 GAATTCGGCGATTCCTTGGAAATTCATTTTGGAAACCAAGTCTAGACGCAACGAAT 628  
 |||||  
 QY 129 ValGluGlnSerPheMetThrMetAlaIleGluIleLysArgMetGlyProGlyAla 148  
 |||||  
 DB 629 GTAGACACTCTTTCATGACGATGCGAGCTGATTAAGAAACGAAATGGGTCTGGACGA 688  
 |||||  
 QY 149 ThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 |||||  
 DB 689 ACAGCTGGTGTGCGACAGAGAAATGCAATGGAATTCAGAGCAGCTCCGCTACAGCAGTCA 748  
 |||||  
 QY 169 GlyGlyGlyCysCys 173  
 |||||  
 DB 749 GGTGAGGTTGCTGC 763

RESULT 5  
 MYPT1 1428 bp mRNA linear ROD 12-SEP-1993  
 LOCUS Mouse mRNA for ras-related Ypt1 protein.  
 DEFINITION Y00094.1 GI:55456  
 VERSION Y00094.1 GI:55456  
 KEYWORDS ras-related protein; Ypt1 gene.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.  
 REFERENCES  
 1 (bases 1 to 1428)  
 Gallwitz, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-NOV-1987) Gallwitz D., Max-Planck-Institut fuer Biophysikalische Chemie, Abt. Molekulare Genetik, Postfach 2841, D-3400 Goettingen, FRG  
 2 (bases 1 to 1428)  
 Haubruck, H., Disela, C., Wagner, P. and Gallwitz, D.  
 TITLE The ras-related Ypt1 protein is an ubiquitous eukaryotic protein: isolation and sequence analysis of mouse cDNA clones highly homologous to the yeast Ypt1 gene  
 JOURNAL EMBO J. 6 (1987) In press  
 FEATURES  
 source Location/Qualifiers  
 1..1428  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_line="F9 and C3H10T1/2"  
 130..747  
 /note="Ypt1 protein (AA 1-205)"  
 /codon\_start=1  
 /protein\_id="CAA68284.1"  
 /db\_xref="GI:55457"  
 /db\_xref="MGD:MG1:97842"  
 /db\_xref="SWISS-PROT:P11476"  
 /translation="MSSNPEDYLFKLLIGDSGVKSCLLRFADPTYESYSTI  
 GVDKIRTELDGKTIKLIQIMPTAGOEPRRTTSSYRAGHIIYVVDQESNNV  
 KQWLEIDRYASENVNKLIVGKCDLTKKVVDTTAKAFACGPIPLETSARKATN  
 VQSPWTMAAEIKRMRGPRATAGAGKSNVAKIQSTPVKQSGGCC"

misc.feature  
 1401..1406  
 /note="pot. polyA signal"  
 polyA\_site  
 1428  
 /note="polyA site"  
 BASE COUNT 410 a 263 c 324 g 431 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,22e-74 Length: 1428  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 32  
 Query Match: 97.07% Indels: 32  
 DB: 10 Gaps: 1

US-09-820-003B-2 (1-173) x MYPT1 (1-1428)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 |||||  
 DB 130 ATGTCACGACGATGAATCCGAATATGATATTTATTCACACTTCTGATGGCGACTCT 189  
 |||||

Oy		21	G1VvAIGLySserCYsLeuLeuEnArPhAlAAsPThrTYrThG1USeTYR	40
Dd		190	GGGGTGGAAAGTCCTCGCCTCTCCCTTGTTGCAGATGATACGTATACGGAAAGCTAC	249
Oy		41	I1SeSThrlIeGIvAlAsPheLYsIIIAArgThrlIeGIvLeuAspGLYsTyfThIle	60
Dd		250	ATACGCAACTGGGTGTGGATTTCACAGATACGACATAAGATTGAATGGAAAAACAATC	309
Oy		61	LysLeuGlInIle-----	64
Dd		310	AACCTTACAGATATGGGACACAGCACAGGCCAGAAAGATTTCGAACAATCACTTCGAGTTAT	369
Oy		65	-----G1USe-Pheasn	68
Dd		370	TACAGAGNAGCCATGGCATCATAGTTGTTATGATATGTACAGATCAGTAGAGCTTCATAT	429
Oy		69	ASnValLYSGlnTPLeuGlnGluIIEAsPARGYrALAsERGluASnVALAsnLYsLeu	88
Dd		430	AACGTTAAAACAGTGGCTGCAGAGATGATATCGCTACGCCATGAAATGTCCAACAACTTG	489
Oy		89	LEnvAlGLAsnLYScYSAsPLeuThThrLYsLYsVALIsAPyrThThThAlAsLYs	108
Dd		490	TTCGTATGGGAACAATAATGTACCTGCCAACAAAAGTAAGTACGTACCAACAAGCAAAG	549
Oy		109	GIUpheAlAsPseSerLeuGLYLeprPhLeuGluTPHrsERAlALysASnAlEThsAn	128
Dd		550	GAATTGGCAATTCCTCTGAAATTCATTITTSYGAAACAGATGTGTMAAGACCAACAGAT	609
Oy		129	VAlGluGlnSerPheMetThrMeTAlAlAGluIIElyLYsArgMecLYPrGGLYAlA	148
Dd		610	GTGAAACAGTCTTTCATGACGATGGCAGCGAGAATTAAGAAAGCAATGGCTCGAGGCT	669
Oy		149	ThrlAgLyGLyAlAGluLYsSerASnVALysIIeGlnSerThProVALysGlnSer	168
Dd		670	ACAGCGGTGGTGGCCGAGAAAGTCCAATGTTAAATCCAGAGCACTCACTCAAGCACTCA	729
Oy		169	GIYGLYGLYCYS Cys 173	
Dd		730	GGTGGAGGCTGCTGC 744	
RESULT 6				
BC000905				
LOCUS	Homo sapiens, RAB1, member RAS oncogene family, clone MGC:5233	1433 bp	mRNA	linear PRI 12-JUL-2001
DEFINITION	IMAGE:2900705, mRNA, complete cds.			
ACCESSION	BC000905			
VERSION	BC000905.2	GI:14705268		
KEYWORDS	MGC.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.			
TITLE	1 (bases 1 to 1433)			
JOURNAL	Strausberg,R. Direct Submission Submitted (16-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	On Jul 12, 2001 this sequence version replaced gi:12654174. Contact: MGC help desk Email: cgapbs-remail.nhl.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E Consortium (LLNL) DNA Sequencing by: Baylor College Of Medicine Human Genome Sequencing Center Center code: BCM-HSC Web Site: http://www.hsc.bcm.tmc.edu/cdna/ Contact: villalon@bcm.tmc.edu Villalon, D.R., Luna, R.A., Hale, S.M., Bulky, S., Lu, X., Garcia,			

A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 4886442.2  
 clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/HLNl at: <http://image.llnl.gov>  
 Series: TRAX plate: 3 Row: k Column: 21

FEATURES	Location/Qualifiers
source	1. .1433

```

/organism="Homo sapiens"
/db.xref="locusID:5861"
/db.xref="taxon:9606"
/clone="MGC:5233 IMAGE:2900705"
/tissue_type="Placenta, ChorionicArctinoma"
/clone_id="NIK.MGC_10"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
179. .796

```

BASE COUNT	432 a	254 c	326 g	421 t
ORIGIN				

Alignment Scores:	
Pred. No.:	1,23e-74
Score:	860.00
Matches:	173
Percent Similarity:	84.33%
Conservative:	0
Best Local Similarity:	84.33%
Mismatches:	0
Query Match:	97.07%
Indels:	32

US-09-820-003B-2 (1-173) x BC000905 (1-1433)

QY	1	MeSerSerMetIaenProGluTyrAspIleuPheLeuLeuIleGlyAspSer	20
Db	179	ATGTCAGAGATGAATCCCGAATATGATATTATTATCAAGTACTTCGATGCGCACCA	238
QY	21	GLYValGlyIleSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr	40
Db	239	GGGGTGGGAAGCTCTCCCTCTCTTAGGTTGCAGATGATCATATTCAGAAAGCTAC	289
QY	41	IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle	60
Db	299	ATCAGACCAATATGGTGGATTTCAAAATTAAGAATATAGATTAGACGGGAAAAACAATC	358
QY	61	LysLeuGlnIle-----	64
Db	359	AAAGCTTCAATATATGGGACACACAGCGCCAGGAAGAATTTCCAGACATCACCTCCAGTTAT	418
QY	65	-----GluSerPheAsn	68
Db	419	TACAGAGAGCCCATGCGCATCATAGTTGTGTATGATGTGCACATCAGGAGTCCCTTCAT	478
QY	69	AsnValLysGlnTrpLeuGlnGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu	88
Db	479	AATGTTAAACAGTGGCTGCACGAAATATAGTATGCTTAAGCCAGGAAATGTCAACAAATTG	538
QY	89	LeuValGlyAsnLysCysAspLeuThrThrCysLysValValAspTyrThrThrAlaLys	108
Db	539	TTGGTAGGGAACAATGTATCTGCACCAACAAAGAAAGTGTAGCTACCAACACGCAAG	598
QY	109	GluPheAlaAspSerLeuGlyIleProPheLeuGlnTrpSerAlaLysAsnAlaThrAsn	128
Db	599	GAATTTGCTCGATCCCTTGGAAATCCGTTTGTGAAACACAGGCTAAAGATGCACGAT	658
QY	129	ValGlnGlnSerPheMetThrMetAlaIleAluIleLysLysArgMetGlyProGlyAla	148

|||||  
Db 659 GTAGAACGCTTTTCATACGATGCGACCTAGATTAAAGCGAATGGTCCGAGACA 718  
Oy 149 ThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
Db 719 ACAGCTGTGTGCTGCGAAGATCCCAATGTAAATTCAGACACTCCAGTCAAGCAGTCA 778  
Oy 169 GlyGlyGlyCysCys 173  
Db 779 GGTGAGAGCTGCTGC 793  
RESULT 7  
LOCUS BC002077 1444 bp mRNA linear ROD 07 AUG-2002  
DEFINITION Mus musculus, RAB1, member RAS oncogene family, clone MGC:6226  
IMAGE:3592802, mRNA, complete cds.  
ACCESSION BC002077  
VERSION BC002077.1 GI:12805232  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1444)  
AUTHORS Strausberg, R.  
JOURNAL Direct Submission  
Submitted (31-JAN-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
REMARK Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdepaxil.stanford.edu](mailto:mcdepaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.  
FEATURES  
SOURCE  
1..1444  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="MGC:6226 IMAGE:3592802"  
/tissue-type="Mammary tumor. Metallothionien-TGF alpha  
model. 10 month old virgin mouse. Taken by biopsy."  
/clone\_id="NCL CGAP Mam1"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
196..813  
/codon\_start=1  
/product="RAB1, member RAS oncogene family"  
/protein\_id="AAH02077.1"  
/db\_xref="GI:12805233"  
/db\_xref="LocusID:5861"  
/translation="MSSMPEVDYLFKILLIGDSGVKSCILLAFADDTYESYIST  
GVDFKRTILDGKTILOIMDPAGGERFRITSSYRGANGITVYDVDDSEFNAY  
KOHQETDRYASENVKLVGNKCDLTTKYVDITAKERFADSLGIFPLETSANAYIN  
VEOSFMTAAEIKRKMPGATAGGAERSNVKIOSTPVKQSGGCGC"  
BASE COUNT 412 a 270 c 342 g 420 t  
Alignment Scores:

Pred. No.: 1.24e-74 Length: 1444  
Score: 860.00 Matches: 173  
Percent Similarity: 84.39% Conservative: 0  
Best Local Similarity: 84.39% Mismatches: 0  
Query Match: 97.07% Indels: 32  
DB: 10 Gaps: 1  
US-09-820-003b-2 (1-173) x BC002077 (1-1444)  
Oy 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
Db 196 ATGTCACAGCATGAATCCCGAATATGATTTATTCANGTTACTTGTGATGGCATTC 255  
Oy 21 GlyAlaGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
Db 256 GGGGTGGAAAGTCCCTGCTCTCTCTAGGTTCAGATGATGATACGTAAGGAAAGCTAC 315  
Oy 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
Db 316 ATCAGCACAAATGGTGTGTGATTTCAAGTACGAACATAAGATTAGATGGGAAACATC 375  
Oy 61 LysLeuGlnIle-----  
Db 376 AAGCTACAGATATGTGGACACAGCAGCGCAAGAAAGATTTCGAACATCATTCCAGTTAT 435  
Oy 65 -----GluSerPheAsn 68  
Db 436 TACAGAGAGCCCATGCGCATCATGATTGTGTATGATGACAGATCAGAGAGTCTTCAAT 495  
Oy 69 AsnValLysGlnTyrPleuGlnGluIleAspArgTyrIleAspArgGluAsnValAsnLysLeu 88  
Db 496 AACGTTAAACAGTGGCTGCAGAGATGATGATGCTACGCCAGTGAATAATGTCACACAGTTG 555  
Oy 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
Db 556 TTGTTAGGAGCAATATGACCTGACCAACAAGAAATAGTACAGTACACAGCAGCAAG 615  
Oy 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
Db 616 GAATTTGCAGATTCCTCTGGAAATTCATTTTGGAAACACAGTGTAAAGACGCAACGAAT 675  
Oy 129 ValGlnGlnSerPheMetThrMetAlaAlaGluIleLysLysAspMetGlyProGlyAla 148  
Db 676 GTAGAACGCTTTTCATACGATGCGACCTAGATTAAAGCGAATGGTCCGAGACT 735  
Oy 149 ThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
Db 736 ACAGCTGTGTGCTGCGAAGATCCCAATGTAAATTCAGACACTCCAGTCAAGCAGTCA 795  
Oy 169 GlyGlyGlyCysCys 173  
Db 796 GGTGAGAGCTGCTGC 810  
RESULT 8  
LOCUS AF226873 2686 bp mRNA linear ROD 13-FEB-2000  
DEFINITION Mus musculus small GTP-binding protein RAB1a mRNA, complete cds.  
ACCESSION AF226873  
VERSION AF226873.1 GI:9693621  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2686)  
AUTHORS Wu, G. and Dorn, G.W. II.  
JOURNAL Mouse RAB1a, member of RAS oncogene family mRNA  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 2686)  
REFERENCE Wu, G. and Dorn, G.W. II.  
AUTHORS Direct Submission  
JOURNAL Submitted (19-JAN-2000) Internal Medicine, University of  
Cincinnati, 231 Bethesda Avenue, Cincinnati, OH 45267-0590, USA





QY 109 GIUPHEALASPSELEUGLYLLEPROPELEUGLUTHRSERALALYSANALATHRASN 128  
 DB 421 GAATTTGAGATTCCTTGGAATTCATTTTGAACACAGCTGTAAGAACAAAANAAT 480  
 QY 129 VALGIUGINSERPHEMETHETHEALALAGLUILELYLSYARMEGLPROGLYALA 148  
 DB 481 GRAGAACACTCTTTCATGACCATGCGACGCGAGATTAAACCGCATGGCTCTGAGACA 540  
 QY 149 THRAGLGLYALAGLULYSERANVALYSLLEGINSERTHPROVALLYSGINSER 168  
 DB 541 ACAGCTGAGGTGCGAGAGATCCATGTTAAATCCAGACACATCCAGCTCAAGACGCTA 600  
 QY 169 GLYGLYGLYCYS 173  
 DB 601 GGTGGAGGCTGCTGC 615

RESULT 10  
 DYGORABI 841 bp mRNA linear VRF 28-APR-1993  
 LOCUS DISCOPYGE ommata GTP-binding protein (o-rabi) mRNA, complete cds.  
 DEFINITION M38393.1 GI:213122  
 ACCESSION M38393.1  
 VERSION GTP-binding protein.  
 KEYWORDS D.ommata electric lobe electromotor neuron, cDNA to mRNA.  
 SOURCE Discopyge ommata  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; Elasmobranchii; Squalae; Hyposqualae; Pristiogryae; Batoidae; Torpediniformes; Narcinoidae; Narcinidae; Discopyge.  
 REFERENCE Ngsee,J.K., Eiferink,L.A. and Schellier,R.H.  
 AUTHORS A family of ras-like GTP-binding proteins expressed in electromotor neurons  
 TITLE Unpublished (1990)  
 JOURNAL Draft entry and computer-readable sequence for [1] kindly submitted by J.K.Ngsee, 10-SEP-1990.  
 COMMENT Dept of Biological Sciences  
 Herrin lab  
 Stanford University  
 Stanford, CA 94305-5020  
 Draft entry and computer-readable sequence for [1] kindly submitted by J.K.Ngsee, 10-SEP-1990.  
 DEPT OF BIOLOGICAL SCIENCES  
 Herrin lab  
 Stanford University  
 Stanford, CA 94305-5020.  
 FEATURES  
 source  
 1..841  
 /organism="Discopyge ommata"  
 /db\_xref="taxon:7785"  
 /cell\_type="electromotor neuron"  
 /tissue\_type="electric lobe"  
 1..841  
 /gene="o-rabi"  
 79..687  
 /gene="o-rabi"  
 /codon\_start=1  
 /product="GTP-binding protein"  
 /protein\_id="AAA49234.1"  
 /db\_xref="GI:213122"  
 /translation="MNPEYDYLKLLIGDSGVKSCLLRPADFTYNEYSITGVV  
 FKIRIELDGKTIKLIQMDTAGOERFRTSSYRGAGIIVYDVDTQESFNPKOM  
 LQEDIRYSENVNKLIVGNKCDLTKKVVDTTKERADSLGIPLETSAKNATVEQ  
 AFTMAAETKRMGPAGTSGSEKSNVNIQSTPKVSGSGGCC"  
 BASE COUNT 269 a 154 c 194 g 224 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9.05e-73 Length: 841  
 Score: 838.00 Matches: 167  
 Percent Similarity: 83.41% Conservative: 4  
 Best Local Similarity: 81.46% Mismatches: 2  
 Query Match: 94.58% Indels: 32

DB: 5 Gaps: 1  
 US-09-820-003b-2 (1-173) x DYGORABI (1-841)

QY 1 MetSerSerMetAsnProglutyrAspPyrLeuPheLysLeuLeuLleGlyAspSer 20  
 DB 70 ATGTCAACCATGAAATCCGAAATATGACTATTATTCAGAGCTGCTGTATGGAATTCAC 129  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspPheThrTyrThrGlySerTyr 40  
 DB 130 GGTGTGGAAATCTTGCTCTCTCTCTGATTGACATGATACATATACAGAAAGTAT 189  
 QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
 DB 190 ATCAGTACAAATGTGTGATTTTAAATCAGAACAAATAGATTAGACGCAAGACATC 249  
 QY 61 LysLeuGlnIle----- 64  
 DB 250 AAACCTCAAAATCTGGACACGCGCTGTCAGAGCGGTTTCGAACAATCACCTCCAGTTAC 309  
 QY 65 -----GlySerPheAsn 68  
 DB 310 TACAGAGGGGCACATGGCATCATAGTGTGATGATGTTACAGACCAGAGATCGTTTAA 369  
 QY 69 AsnValLysGlnTyrPheGlnGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
 DB 370 AATGTAAACCATGTGCTTCAGGAATATGATGCTTATGTCAGTGAATATGTTAACAAAGTTA 429  
 QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
 DB 430 CTGGTGGGAAACAAATGTGATCTGCACAAACAAAGAGGTGGATTTATACAAACGCAAG 489  
 QY 109 GIUPHEALASPSELEUGLYLLEPROPELEUGLUTHRSERALALYSANALATHRASN 128  
 DB 490 GAATTTGAGATTCCTTGGAATTCATTTTGAACACAGTGCAAAATGCAACAAAT 549  
 QY 129 VALGIUGINSERPHEMETHETHEALALAGLUILELYLSYARMEGLPROGLYALA 148  
 DB 550 GTAGAACCAAGCTTCATGACTATGCTGCTGATTAACAAACGAATGGCTCGGAGCT 609  
 QY 149 THRAGLGLYALAGLULYSERANVALYSLLEGINSERTHPROVALLYSGINSER 168  
 DB 610 ACATCTGTGTGTTCTGGAAGAAGTAAATATATATATACCAACGACCGCATAGTCGCT 669  
 QY 169 GLYGLYGLYCYS 173  
 DB 670 GGTGGGAGGCTGCTGC 684

RESULT 11  
 AR070365 925 bp DNA linear PAT 18-FEB-2000  
 LOCUS Sequence 4 from patent US 5892012.  
 DEFINITION AR070365  
 ACCESSION AR070365  
 VERSION AR070365.1 GI:7221253  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 925)  
 AUTHORS Hillman,J.L., Lal,P., Corley,N.C. and Shah,P.  
 TITLE Rab proteins  
 JOURNAL Patent: US 5892012-A 4 06-APR-1999;  
 FEATURES  
 source  
 1..925  
 /organism="unknown"  
 BASE COUNT 198 a 263 c 288 g 176 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5.92e-66 Length: 925  
 Score: 768.50 Matches: 156  
 Percent Similarity: 80.39% Conservative: 8  
 Best Local Similarity: 76.47% Mismatches: 7

[illegible]

Pred. No.:	1,46e-65	Length:	1995
Score:	768.50	Matches:	156
Percent Similarity:	80.39%	Conservative:	7
Best local Similarity:	76.47%	Mismatches:	8
Query Match:	86.74%	Indels:	33
DB:	6	Gaps:	2

  

US-09-820-003B-2 (1-173) x AX086155 (1-1985)	
QY 2 SerSerMetAsnProGluTyrAspTyrLeuPheIysLeuLeuIleGlyAspSergIy 21	
DB 42 GCGCCATGAACCCCGAATATACCTTTTAAGCGCTTTGTGATGGCACTGAGC 101	
QY 22 ValGlyIysSerCysLeuLeuLeuAlaPheAlaAspThrTyrThrGluSerTyrIle 41	
DB 102 GTGGGCAAGATATCCCTGGCTCCTGGGCTTTGCTGATGACAGTACAGAGAGCTGCATAC 161	
QY 42 SerThrIleGlyValAlaPheIysIleArgThrIleGluLeuAspGlyIysThrIleIys 61	
DB 162 AGGACCATGGGGGTGGACTTCATAGATCCGAAACCATCGAGCTGATGGCAAAATCAATA 221	
QY 62 LengJinIle----- 64	
DB 222 CTCACAGATCTGGAGACACAGCGGCGAGAACGGTTCGGACATCACTTCAGCTACTAC 281	
QY 65 -----GluSerPheAsnAsn 69	
DB 282 CGGGGGGCTCATGCGATCTGTGGTATGACGTCACCTGACGAGCAAGATCTCCAGCCAAC 341	
QY 70 ValIysGlnTyrPheIysGlnIleAspArgTyrAlaSerGluAsnValAlaAspIysLeu 89	
DB 342 GTGAGCATGATGGCTCGAGGAGATATGACGGTATGACGACGAGAACCTCAATTAACCTCTG 401	
QY 90 ValGlyAsnIysCysAspIleuThrThrIysIysValValAspTyrThrThrAlaIysGlu 109	
DB 402 GTGGGCAACAGACCACTCTACACCAAGAGTGGTGGACAAACCAACCAAGCAAGAG 461	
QY 110 PheAlaAspSerIleuGlyIleProPheLeuGluThrSerAlaIysAsnAlaThrAsnVal 129	
DB 462 TTTCGCAATCTCTGGGCAATCCCTCTCTTGGAGAGACAGCGCAAGATCCCAACATGTC 521	
QY 130 GluGlnSerPheMetThrMetAlaIleGluIleIysIysArgMetGlyProGlyAlaThr 149	
DB 522 GACACAGCGCTTCATGACCATGTGCTCTGGAATATAAAAGCGATGGGGGCTGGAGAGCC 581	
QY 150 AlaGlyIysAlaGluIysSerAsnValIysIleGlnSerThrProValIysGlnSergIy 169	
DB 582 TCGGGGGGGG---GAGCGGGCCAAATCTCAAGTCGACAGCACCCCTGTAAAGCGCGCTGCG 638	
QY 170 GlyIysCysCys 173	
DB 639 GGTGGCTGTGTC 650	

  

RESULT 13	
ISM801608	1985 bp mRNA linear PRI 20-MAR-2002
LOCUS	Homo sapiens mRNA; cDNA DKFZ564I172 (from clone DKFZp564I172);
DEFINITION	complete cds.
ACCESSION	AL136635
VERSION	AL136635.1 GI:12052795
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 1985)
JOURNAL	Duesterhoeft,A., Lauber,J., Mewes,H.W., Gaassenhuber,J. and Wiemann,S.
COMMENT	Direct Submission Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;



```

OY 2 SerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSergly 21
DB 6 GCCCCATGACACCCGATATGACTACCTGTTAAAGCGCTTTGATGTGAGTACCTGGGG 65
OY 22 ValGlyLysSerCysLeuLeuLeuAspPheAlaAspPheThrGlyGlySerTyrIle 41
DB 66 GTGGGCAAGTCATGCTCTCTCTGCGGTTTGTCTATGACACGACGACAGAGGTATATC 125
OY 42 SerThrIleGlyValAspPheLysIleArgThrIleGlyLeuAspGlyLysThrIleLys 61
DB 126 AGCACCTTGGGCTGGACTTCATGATTCGACCATTCGATGCGATGGCAAAACATCATA 185
OY 62 LeuGluIle----- 64
DB 186 CTACAGATTGGGACACAGCTGTGTCAGAAACGTTCAAGACACTCAGCTACTACTAT 245
OY 65 -----GluSerPheAsnAsn 69
DB 246 CGGGGTCTCAGCAGCATCTGTTGTTATGACGTACATGACGACGATCTACCTACTAT 305
OY 70 ValLysGluThrLeuGluGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuLeu 89
DB 306 GTGAAGACAGTGGCTGCAGAAATAGATCGCTACGCCAGTGAATGATCAATTAACCTG 365
OY 90 ValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrAlaLysGlu 109
DB 366 GTACGCAACAGAGTACCTTACCACCAAGAGTCTGTGACATTCACGACCCAGAGGA 425
OY 110 PheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129
DB 426 TTTCACACACTCTCTGGGTGCTCCCTTCCTGCGAGACAGATCCAAAGATCCCAATG 485
OY 130 GluIleAspPheMetThrMetAlaAlaGluIleLysLysArgMetCylProGlyAlaThr 149
DB 486 GAAAGGCAATATATCAATAGCTGTCAGAGATCAAAAGAGATGGGCGCAGGACAGCA 545
OY 150 AlaGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSergly 169
DB 546 TCTGGGGGCT---GAAGGCCCAACCTGAGATGACGACACACTCTGTGAAATCTGTAG 602
OY 170 GlyGlyCys 173
DB 603 GGTGGCTGCTGC 614

RESULT 15
AC106368/c 161955 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-117N16, *** SEQUENCING IN PROGRESS
DEFINITION *** 71 unordered pieces.
ACCESSION AC106368
VERSION AC106368.2 GI:21731746
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
          Chordata; Craniata; Vertebrata; Euteleostomi;
          Eukaryota; Metazoa; Chordata; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
REFERENCE 1 (bases 1 to 161955)
AUTHORS Muzny,D.N., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
        Albrock,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
        Barbata,J., Benton,J., Blinze,K., Blankenburg,K., Bonnin,D.,
        Bouck,J., Bowie,S., Britten,M., Brown,E., Brown,N.P.,
        Bunay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
        Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
        Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
        Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
        Davila,M.L., Davis,C., Davy-Carroll,L., Dedetich,D.A.,
        Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
        Douthett,K.J., Draper,H., Dugan-Rocha,S., Durkin,K.J.,
        Eathart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
        Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
        Gabist,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gall,R.,

```

```

CORRELL,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
HARRIS,C., HARRIS,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kravits,E., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louise,H.,
Lozano,R.J., Lu,X., Luetter,A., Luetter,R., Luna,R., Ma,J.,
Manshwar,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,N., Morris,S.,
Moser,N., Neal,D., Newton,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,
Oragunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swalek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,C., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,Y., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 161955)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 161955)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:18138895.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web Site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center Project name: GKTO
Center clone name: CH230-117N16
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 102133 bases at least Q40
Consensus quality: 108299 bases at least Q30
Consensus quality: 114299 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 71 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 1109 1108: contig of 1108 bp in length
* 1209 1208: gap of unknown length
* 1209 2485: contig of 1277 bp in length
* 2486 2585: gap of unknown length
* 2586 3731: contig of 1146 bp in length
* 3732 3831: gap of unknown length

```

```

* 3832 5080: contig of 1249 bp in length
* 5081 5180: gap of unknown length
* 5181 6572: contig of 1392 bp in length
* 6573 6673: gap of unknown length
* 6673 8117: contig of 1445 bp in length
* 8118 8217: gap of unknown length
* 8218 9450: contig of 123 bp in length
* 9451 9551: gap of unknown length
* 9551 10589: contig of 1039 bp in length
* 10590 10689: gap of unknown length
* 10690 12382: contig of 1593 bp in length
* 12383 12382: gap of unknown length
* 12383 13857: contig of 1475 bp in length
* 13858 13957: gap of unknown length
* 13958 14990: contig of 1033 bp in length
* 14991 15090: gap of unknown length
* 15091 16216: contig of 1126 bp in length
* 16217 16316: gap of unknown length
* 16317 17813: contig of 1437 bp in length
* 17814 17913: gap of unknown length
* 17914 19393: contig of 1480 bp in length
* 19394 19493: gap of unknown length
* 19494 20579: contig of 1086 bp in length
* 20580 20679: gap of unknown length
* 20680 22114: contig of 1435 bp in length
* 22115 22214: gap of unknown length
* 22215 23397: contig of 1083 bp in length
* 23398 23397: gap of unknown length
* 23398 25553: contig of 2156 bp in length
* 25554 25653: gap of unknown length
* 25654 27071: contig of 1418 bp in length
* 27072 27171: gap of unknown length
* 27172 28314: contig of 1143 bp in length
* 28315 28414: gap of unknown length
* 28415 29996: contig of 1582 bp in length
* 29997 30096: gap of unknown length
* 30097 31329: contig of 1233 bp in length
* 31330 31429: gap of unknown length
* 31430 32584: contig of 1155 bp in length
* 32585 32684: gap of unknown length
* 32685 34267: contig of 1583 bp in length
* 34268 34367: gap of unknown length
* 34368 36135: contig of 1768 bp in length
* 36136 36235: gap of unknown length
* 36236 37482: contig of 1247 bp in length
* 37483 37582: gap of unknown length
* 37583 39436: contig of 1854 bp in length
* 39437 39536: gap of unknown length
* 39537 40785: contig of 1249 bp in length
* 40786 40885: gap of unknown length
* 40886 42446: contig of 1561 bp in length
* 42447 42546: gap of unknown length
* 42547 45236: contig of 2690 bp in length
* 45237 45336: gap of unknown length
* 45337 47003: contig of 1667 bp in length
* 47004 47103: gap of unknown length
* 47104 48616: contig of 1513 bp in length
* 48617 48716: gap of unknown length
* 48717 50883: contig of 2167 bp in length
* 50884 50983: gap of unknown length
* 50984 53266: contig of 2283 bp in length
* 53267 53366: gap of unknown length
* 53367 54563: contig of 1197 bp in length
* 54564 54663: gap of unknown length
* 54664 56471: contig of 1808 bp in length
* 56472 56571: gap of unknown length
* 56572 58274: contig of 1703 bp in length
* 58275 58374: gap of unknown length
* 58375 60721: contig of 2347 bp in length
* 60722 60821: gap of unknown length
* 60822 61918: contig of 1097 bp in length
* 61919 62019: gap of unknown length
* 62019 64331: contig of 2313 bp in length

```

```

* 64332 64431: gap of unknown length
* 64432 66337: contig of 1806 bp in length
* 66338 66337: gap of unknown length
* 66338 69384: contig of 3046 bp in length
* 69384 69384: gap of unknown length
* 69384 71719: contig of 2236 bp in length
* 71720 71820: gap of unknown length
* 71820 73999: contig of 2180 bp in length
* 74000 74099: gap of unknown length
* 74100 76398: contig of 2299 bp in length
* 76399 76498: gap of unknown length
* 76499 78819: contig of 2321 bp in length
* 78820 78919: gap of unknown length
* 78920 82063: contig of 3144 bp in length
* 82064 82163: gap of unknown length
* 82164 84718: contig of 2555 bp in length
* 84719 84818: gap of unknown length
* 84819 86848: contig of 2030 bp in length
* 86849 86948: gap of unknown length
* 86949 89201: contig of 2253 bp in length
* 89202 89301: gap of unknown length
* 89302 92038: contig of 2737 bp in length
* 92039 92138: gap of unknown length
* 92139 95048: contig of 2910 bp in length
* 95049 95148: gap of unknown length
* 95149 98066: contig of 2918 bp in length
* 98067 98166: gap of unknown length
* 98167 100925: contig of 2759 bp in length

```

## Alignment Scores:

```

Pred. No.: 1,02e-62 Length: 161955
Score: 762.50 Matches: 154
Percent Similarity: 79.90% Conservative: 9
Best Local Similarity: 75.49% Mismatches: 8
Query Match: 86.06% Indels: 33
DB: 2 Gaps: 2

```

US-09-820-003B-2 (1-173) x AC106368 (1-161955)

```

QY 2 SerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSergly 21
   :::::::::::
Db 161034 GCCGCATGACACCCCGCAATATACCTACCTGTTAACTGCTTTTATTTGGTACCGGCC 160975
QY 22 ValGlyLysSerCysLeuLeuLeuArgPheAlaAspPheTyrThrGluSerTyrIle 41
   |||||||
Db 160974 GTGGCAAGTCATGCTGCTGCTGCGGTTGTGTGATGACACGTACACAGAGAGTACATTC 160915
QY 42 SerPheIleGlyValAlaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61
   |||||||
Db 160914 AGCACCATTTGGGTGAGTCTCAAGATTCGAACATTGAACGTGAGCAAAACCATCAAA 160855
QY 62 LeuGlnIle-----GluSerPheAsnAsn 64
   |||||||
Db 160854 CTACAGATTTGGGACACAGCTGCTCAGGACGTTCCAGACCATCTACCTACTAT 160795
QY 65 -----GluSerPheAsnAsn 69
   |||||||
Db 160794 CGGGGTCTCATGCGCATCATTTGGTGTATGACGTCACTGACAGCAAGAAATCTACCTAAT 160735
QY 70 ValLysGlnTPrLeuGlnGluIleAspArgTyrAlaSerGluAsnValAlaLysLeuLeu 89
   |||||||
Db 160734 GTGAACAGTGGCTCAGAAATAGATCGCTACGCGCAGTGAATGCAATAAACTGCTG 160675
QY 90 ValGlyAsnLysCysAspLeuThrThrLysLysValValaAspTyrThrAlaLysGlu 109
   |||||||
Db 160674 GTAGCAACAGAGTGGCTCAGCAACCAAGAGTGTGGAGCAATATACAGCAAGCAAGAA 160615
QY 110 PheAlaAspSerLeuGlyTyleProPheLeuGlnTPrSerAlaLysAsnAlaThrAsnVal 129
   |||||||
Db 160614 TTTCAGACTCTCTGGGTGTCCCTTCTGTGAGACAAATGCCAAGAAATGCCCAATGT 160555
QY 130 GluInSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr 149
   |||||||

```



